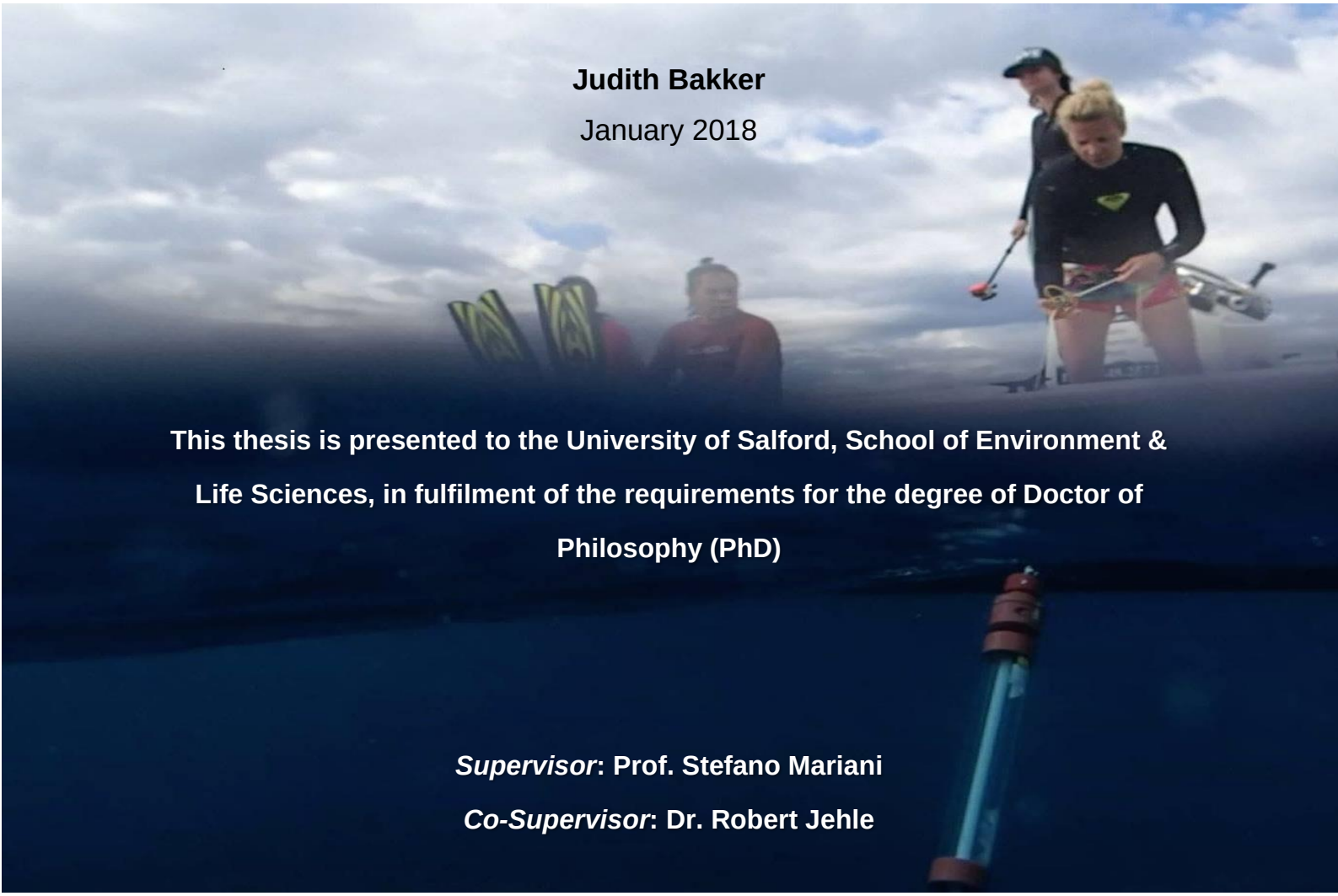
e(lasmo)DNA:

The role of environmental DNA (eDNA) analysis in marine fish biodiversity assessment, with special focus on elasmobranchs



Judith Bakker

January 2018

This thesis is presented to the University of Salford, School of Environment & Life Sciences, in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

***Supervisor*: Prof. Stefano Mariani**

***Co-Supervisor*: Dr. Robert Jehle**

University of Salford MANCHESTER

THE PEW CHARITABLE TRUSTS

Table of Contents

List of figures

List of tables

Acknowledgements

Dear **Stefano**, who would have thought we would be here, when I contacted you four years ago to kind of randomly inquire whether you were interested in this shark PhD project idea that I had put together. During our first Skype call you said, 'well yes, but how about instead, we apply for a grant to use eDNA on sharks?'. And I had to admit, that at that point, I had never even heard of eDNA. You mentioned that you had a contact at Pew and that you were going to see if they were interested in potentially funding this project. After a few days you told me to put a small proposal together for the next day (!) because it turned out they actually were interested. So, I spent the whole night on the first draft of a proposal, and another few days later we had the first Skype meeting with Pew to discuss the project. A couple of days and proposal drafts further on, and they agreed to fund us for the first half of the project, and all of a sudden, I had a PhD and moved to Salford. And what an adventure it has been. We have been to many conferences and on field work to Belize together, we even made a documentary (GoPro GoPro!!). There have also been some trying times, such as when half-way through, we had to find additional funding for me to be able to continue the project and finish the PhD. And even though there have been moments when we wanted to kill each other, it has been pretty amazing. I am thankful for your support, encouragement and guidance throughout, but I also appreciate the freedom you have given me to develop as a researcher and to do certain things my own way, even though you didn't always approve and sometimes it took me a while to appreciate your take on things (which happens when you put two hard-headed people together). I am glad that I got to make this PhD journey with you as my supervisor and look forward to our future collaborations.

Dear **Owen**, I am still convinced that hiring you as a post-doc was one of Stefano's best ideas during my time in Salford. Not only are you a brilliant scientist, but also one of the kindest people I know. I have learned incredibly much from you, and you have been an absolutely crucial factor in the progression of my PhD, but also in my development as a researcher. I wish you and Sandra all the best (all the way up) in Tromsø and I am glad that our collaboration will continue.

I owe thanks to **Andy**, who has been a crucial person at the beginning of my PhD. I know you had some doubts at first, as I really didn't have much molecular lab experience, but I am glad I managed to quickly change your mind and that you were there to teach me the ropes.

I am very grateful to Pew for believing in us and the project, and in particular to **Katie** and **Angelo (**we even got to snorkel with hammerheads together!), who were instrumental in securing the funding, and even though they both moved on to different jobs/projects, they have been supportive right until the finish. Thank you also to **Becky**, **Dom** and **Anthony**, for the long breath and continued support, until we finally got the first manuscript published.

I am also very grateful to 'the University of Salford R&E strategy funding' and in particular to **Janet** and **Marie** for their support, and who half-way through the project, without much hesitation, agreed to take me on as an iCase student, providing us with the funding and support to continue and finish my PhD.

Many thanks to all our collaborators, who I certainly hope to keep working with in the future, for their support and expertise during the sampling campaigns and manuscript preparations. You all played a vital role in the project and without you, my PhD would have looked very different; **Demian**, **Dayne**, **Tristan**, **Heidi**, **Germain**, **David** and **Laurent**. A special thanks to **Austin**, who contacted us for a collaborative project and subsequently took me along to the British Virgin Islands for sampling (and an overall awesome week with **Sami** and **Diejo** and everyone involved with the **BVI Art Reef** project).

Sara, ik moet nog altijd denken aan onze eerste ontmoeting, jij en Stefano kwamen ons kantoor binnen terwijl ik mn nagels aan het lakken was, dat had een beetje een rare indruk op je achter gelaten. Maar gelukkig bleek al snel dat we het ontzettend goed met elkaar konden vinden (en dat was zeker niet vanwege het feit dat we beiden zoveel klaagden over Salford..... ;). Tijdens het jaar dat je in Manchester was, en in dezelfde flat kwam wonen, hebben we een vreselijk goede tijd gehad en ik ben blij dat we nog steeds zo goed bevriend zijn.

Together with Sara and me, **Peter** was part of 'Team Awesome'. You were the only other PhD student with Stefano at the time I started, and you were essential in shaping my first period in Salford, so when you finished your PhD and moved to Hull, it was really the end of an era. I am glad you have found your place in California and hope we will get to work on some awesome projects together.

I am also grateful to **Cristina** (together with Sara, part of team 'No whining') and **Stella**, for both your friendship, and help and expertise in the lab on various projects, you two made a lot of lab work a lot less boring.

D and Kippie Kapokkie, **Donna** and **Naiara**, where to start? We have been through many (academic and personal) ups and downs together during our shared time in Salford, which has made the downs a lot more bearable and the ups so much more enjoyable. I am very grateful for our friendship. I have immensely enjoyed all of our times and adventures together, and in particular our time in Afffffffrika. I am looking forward to our next travels.

Cristina, **Nico** and **Matthew**, without you there would have been no 'Traces'. But I am also grateful for your friendship, and our time together in Bimini was an absolute pleasure.

A lot of troubles and laughter have been shared with my office mates and friends, **Andjin (**ik ben blij dat je gebleven bent en we hebben het allebei toch maar mooi gehaald!), **Ross**, **Moo**, **Charlie**, **Becky** and **Donal**, but also **Charles,** even though still residing in Cockcroft. Thank you, for being there.

In addition to my colleagues, there is another group of people I want to thank for providing support and distraction during my PhD, and without whom I would never have survived the past four years.

Papa en **Mama**, zonder jullie was dit natuurlijk nooit gelukt. Jullie hebben me altijd gesteund in alles, ook al waren jullie het misschien niet altijd eens met waar ik nou weer mee bezig was (of waar ik nu weer heen op reis ging), en ben ik zeker niet altijd even makkelijk geweest. Maar ondanks dat hebben jullie me altijd de vrijheid gegeven mijn eigen weg te vinden, al was dat niet altijd de meest directe of makkelijkste weg, maar dit heb ik wel bereikt dankzij jullie. Ook jullie, **Iris**, **Steph**, **Aeron** en '**Baby**', zijn een cruciaal onderdeel van mijn basis.

Vanessa, **Roy**, **Rinke**, **Henna**, **Yennick** en **Revée**, vanaf het moment dat Vanessa en ik bevriend zijn geraakt, hebben jullie me opgenomen in de familie. Ik ben altijd welkom in de 5 sterren Hoppe accommodatie, en tijdens mijn PhD heb ik echt een aantal ontsnappings weken bij jullie doorgebracht, bedankt voor al jullie steun.

Jonas, **Ylva**, **Nasim**, **Sabina** and **Mattias** and **The Poepie**, you have always been like family to me and I am very grateful for your friendship and support, and for always making me feel welcome.

Thanks to my (Amsterdam and beyond) home base, without your support, knowing you were there, this would have been so much harder; **Vanessa**, **Merredith** en **Jeroen**, **Linda**, **Marjan**, **Ferial**, **Charlotte** (en de 'cute aunties'). **Mai**(-Britt), **Camillidae** (Camilla) and **AM** (Annemarie). **Christine** en **Sarah** en onze drie-eenheid op afstand, ik zie uit naar de rituele thesis verbranding. En ook; **Coen** en **Felisa**, **Sanne** en **Menno**, **Julia**, **Thomas**, **Jesse**. **Marlies**, **Femke** en **Etienne**. **Juul** en **Milo**, **Eelke** en **Pascal**.

And my Mancunian based friends who provided me with much needed motivation and distraction; like-minded co-traveller and detour-taker **Frederica** and my 'gym & drinks' buddies **Margherita** and **Jess**.

And finally, thank you **Phil**, for being the awesomest, and especially for your unrelenting support and encouragement, even from a distance, you make everything better.

Abstract

Knowledge of spatial and temporal variation in abundance is critical for the implementation of effective protective measures for organisms that are both naturally rare and vulnerable to exploitation. Therefore, the development of management and conservation strategies for taxa like teleosts and elasmobranchs, depends on the accurate assessment and monitoring of the distribution and abundance of target species. However, detecting species occurrences is often even more challenging in the aquatic environment than on land. Consequently, as is the case for many mobile, and often rare, vertebrates, fish (and particularly shark) detection is inherently difficult. Environmental DNA metabarcoding, based on the retrieval of genetic traces (skin cells, metabolic waste, etc.) naturally released in the environment, is emerging as a non-invasive method for the detection and identification of rare and elusive species in a wide range of ecosystems, including aquatic environments. My thesis addresses the development and application of an environmental DNA (eDNA) approach for the assessment of marine communities, and particularly of elasmobranch species. This novel eDNA approach was developed to investigate elasmobranch diversity in order to assess species richness in areas of special conservation concern. While simultaneously examining the influence of interacting factors such as habitat type and conservation regime in determining diversity and abundance. Additionally, the performance of eDNA analysis was compared with more traditional sampling methods. Moreover, the performances of multiple markers for the detection and characterization of both elasmobranch and teleost diversity were tested and evaluated. The potential implications of eDNA for fish, and larger scale marine community assessment and monitoring, spatial planning and fisheries management are significant.

1 Chapter I

Introduction and Background

1.1 What is environmental DNA (eDNA)?

All organisms continuously leave traces of themselves behind in the environment, in the form of shed skin cells, hairs, bodily fluids, metabolic waste, gametes or blood. Any of these materials can contain pieces of the organism's DNA. Environmental DNA (eDNA) analysis is based on the retrieval of this naturally released genetic material from the environment. It generally refers to bulk DNA extracted from an environmental sample such as water, but also from soil, sediment, snow, or even from air (Taberlet et al. 2012). In aquatic systems, macro-organismal derived eDNA can be present as free DNA, cellular debris or particle bound DNA, and is mostly present in small fragments, owing to rapid degradation (Turner, Barnes, et al. 2014). But, much of the eDNA is retrieved from cellular material and may therefore contain still relatively undamaged nucleic acid molecules. Nevertheless, eDNA studies mostly focus on the detection of short fragments, as currently available parallel sequencing and qPCR platforms have short-read capabilities limited to a few hundred base pairs. When DNA is present at low concentrations, mitochondrial DNA (mtDNA) is often targeted, since there are substantially more mitochondrial than nuclear DNA copies per cell (Wilcox et al. 2013). Commonly employed mtDNA genes include cytochrome b, cytochrome c oxidase 1 (COI), 12S rRNA, and 16S rRNA (Kelly, J. a. Port, et al. 2014; Thomsen, Kielgast, Lars Lønsmann Iversen, et al. 2012; Valentini et al. 2016), and targeted fragments typically fall within the range of 79-285 bp (Ficetola *et al.*, 2008; Minamoto *et al.*, 2012). The level of target specificity is often the main determining factor when choosing primers for eDNA analysis. Environmental DNA is emerging as a non-invasive method for the detection and identification of rare and elusive species in a wide range of ecosystems, including aquatic environments (Port et al. 2016; Thomsen, Kielgast, Lars Lønsmann Iversen, et al. 2012; Yamamoto et al. 2017). It is rapidly diffused from its source and degraded under the influence of local environmental conditions such as mechanical forces, UV radiation, pH, temperature (Barnes *et al.*, 2014; Jerde *et al.*, 2011;Pilliod *et al.*, 2014), microbial activity (Barnes et al. 2014a), and spontaneous chemical reactions such as oxygenation (Nielsen et al.

2007; Lindahl 1993). This indicates a low probability of long-distance dispersal of eDNA in aquatic ecosystems (Thomsen and Willerslev, 2015). Hence, the detection of eDNA from a specific taxon indicates its presence or very recent presence in the environment (O'Donnell et al. 2017). However, there remains much uncertainty on the impact of oceanic currents on the local-scale spatial patterns of trace DNA, especially in open marine systems (Hajibabaei et al. 2006; Taberlet et al. 2012).

Owing to recent advances in high-throughput sequencing and bioinformatics, the use of eDNA has developed into a cost-effective and rapid, non-invasive method for collecting and analysing biological samples from large portions of the environment without the necessity of isolating the target species (Handelsman 2005). Using this approach, thousands of species present in any environmental sample can be detected by high-throughput DNA sequencing and identified using molecular taxonomic databases, thus revolutionizing our ability to detect species and conduct genetic analysis for conservation, management and research of aquatic ecosystems.

1.2 A short history of eDNA

Over the past decade, there has been substantial development of eDNA recovery and sequencing techniques, which have resulted in an increasing interest in its use as a tool for both targeted species detection and biodiversity assessments (Handelsman, 2005). The term 'eDNA' was first used by microbiologists, who have been applying the eDNA method since the mid-1980's, to assess the diversity of micro-organism communities in ancient marine sediments (Willerslev et al. 2003). The general eDNA methods currently used for monitoring aquatic populations arose from this early work (Bailiff & Karl 1991; Paul et al. 1996; Weinbauer et al. 1993). Subsequently, in the 1990's, eDNA methods were employed to monitor phytoplankton blooms and to assess changes in biomass of bacterial and viral communities (Thomsen & Willerslev 2015). The use of eDNA has more recently been developed to elucidate macro-organism identity in aquatic environments. However, the nature of eDNA from macro-organisms in environmental samples is different from that of microbial organisms (prokaryotes and microbial eukaryotes) because the former are present only as part of the organism (cellular remains or free DNA), whereas the latter may be detected by DNA derived from whole, living organisms present in the samples (Thomsen and Willerslev, 2015).

Environmental DNA as a method to assess the diversity of macro-organismal communities was first applied to sediments, revealing DNA from extinct and extant mammals, birds and plants (Ficetola *et al.*, 2008). In 2008, the eDNA method was applied for the first time to confirm the presence of an aquatic invasive species, the American bullfrog (*Rana catesbiana*), from water samples in a natural lentic system (Jerde et al. 2011). Subsequently, the first eDNA study in freshwater lotic systems for the detection of invasive Asian carp was published in 2011 (Foote et al. 2012). In 2012, eDNA analysis was first applied to the marine environment for the detection of marine mammals (Thomsen, Kielgast, Lars Lønsmann Iversen, et al. 2012), and for the estimation of marine fish diversity (Piaggio et al. 2014; Pilliod et al. 2014; Yamamoto et al. 2017). Environmental DNA has since been applied for the detection of a large range of aquatic species in both freshwater and marine systems (Sigsgaard *et al.*, 2016; Simpfendorfer *et al.*, 2016; Gargan *et al.*, 2017; Weltz *et al.*, 2017), and more recently, for the detection of sharks and rays (Sigsgaard *et al.*, 2016; Simpfendorfer *et al.*, 2016; Bakker *et al.*, 2017; Gargan *et al.*, 2017; Weltz *et al.*, 2017).

1.3 eDNA vs. traditional monitoring techniques

Currently established survey methods, such as fishing by long-lining or gill-netting, acoustic or satellite tagging and monitoring, baited remote underwater video (BRUV), underwater visual census (UVC), ecological knowledge surveys and fisheries-dependent population surveys, all have associated biases and challenges. These include being potentially resource intensive, selective and dependent on taxonomic expertise, and sometimes invasive and potentially traumatogenic (Lodge *et al.*, 2012; Simpfendorfer *et al.*, 2016; Wheeler, 2004). Traditional surveys are also highly susceptible to false negatives; failing to detect rare of cryptic species that are present. Therefore, the assessment and monitoring of the distribution and abundance of mobile species in aquatic environments remains challenging and would benefit from new, complementary methods of investigation.

Environmental DNA has been shown to be a reliable detection method, matching, or even outperforming, conventional survey methods (Dejean et al. 2011; Huver et al. 2015). Since eDNA analysis is an inherently non-invasive detection method, there is no necessity for the species of interest (or its habitat) to be either disturbed or caught, in order to establish their presence, or to acquire a positive taxonomic identification. Without the need for visual detection, using eDNA makes it easier to detect rare species, or those species which have juvenile stages that closely resemble other species (Dejean *et al.*, 2011; Huver *et al.*, 2015).

Species of conservation concern often have low population numbers, making surveys based on eDNA methods particularly suitable for informing applied conservation efforts (Foote *et al.*, 2012; Olson *et al*., 2012; Thomsen *et al.*, 2012). Likewise, exotic and invasive species are typically rare at their expanding range margins, requiring highly sensitive detection methods (Dejean et al. 2011). False negatives in presence-absence data using traditional methods can prevent effective habitat protection for threatened species. A particular case study concerning shark species has recently been described for the New Caledonian archipelago. Here, both 2,758 Under Water Visual Censuses (UVC) and 385 Baited Underwater Video Stations (BRUVs) detected 9 shark species. While with only 22 eDNA samples, 13 shark species were detected. Thus, despite two orders of magnitude less sampling effort, with eDNA analysis, 44% more shark species were detected compared to UVC's and BRUV's, revealing a greater diversity of sharks than previously thought. Hence, indicating the need for large-scale eDNA assessments to improve shark monitoring and conservation efforts (Boussarie *et al*. 2017 in press).

In terms of sampling effort, eDNA analysis can offer considerable time and cost benefits (Rees et al. 2014; Valentini et al. 2016), especially concerning the distribution of rare and threatened species (Simpfendorfer et al. 2016). Collecting water samples requires significantly less time and resources compared to traditional survey methods. This is particularly true when target species are found in remote and/or difficult to access areas (Jerde et al. 2011; Rees et al. 2014). During a study of invasive Asian carp detection, it took 93 days of person effort to detect one silver carp (*Hypophthalmichthys molitrix*) by electrofishing, whereas eDNA analysis required only 0.174 days person effort to achieve a positive detection (Giles et al. 1980). Additional advantages of the eDNA method compared to traditional sampling relate to the ease of obtaining permits for the collection and handling of water samples *vs*. (live) animals. And sampling can often be carried out under more extreme weather conditions

There are however, important caveats associated with eDNA detection and traditional survey methods still have a number of advantages over eDNA methodologies. Foremost, using eDNA analysis, it is not possible to distinguish whether the detected DNA from a certain species has been released by a dead or live animal. However, the analysis of environmental RNA (eRNA) is an emerging technique, which may help overcome this issue (Pochon et al. 2017). Additionally, eDNA methods do not provide information on size, movement patters, condition, developmental stage (eggs, larvae, juveniles, adults) and sex of the target organism. Moreover, the detection of an individual's DNA, without direct observation, cannot provide information on the exact location of the animal. Furthermore, when using a mitochondrial marker, of which the DNA is mostly maternally inherited (Mahon et al. 2013; Eichmiller et al. 2014), it will not be possible to distinguish hybrids (which may be the result of breeding between native and invasive species), from their maternal species. Lastly, inferring abundance information from eDNA remains challenging and is a key area for further research (Doi et al. 2017; Klobucar et al. 2017; Doi, Uchii, et al. 2015).

Environmental DNA is becoming a rapid and cost-effective tool for collecting species' presence, distribution, and with some caveats, (relative) abundance data. Most likely, with continuing development in the fields of DNA sequencing and bioinformatics, eDNA methods will increasingly complement (rather than completely replace) traditional survey methods.

1.4 eDNA approaches: species-specific vs. metabarcoding

Currently, the use of eDNA can broadly be divided into two main approaches, a single-species approach (eDNA barcoding) and a multi-species approach (eDNA metabarcoding). Environmental DNA barcoding is aimed at detecting a single species in the environment by Polymerase Chain Reaction (PCR) or quantitative PCR (qPCR). Whether the amplified eDNA sequences belong to the target species, is then often confirmed through Sanger sequencing (Mahon *et al.*, 2013; Eichmiller, Bajer and Sorensen, 2014). Conventional PCR has previously been used for species-specific eDNA detection (Bourlat et al. 2013). However, quantitative PCR (qPCR) offers a distinct advantage over traditional endpoint PCR techniques through the addition of a fluorescent dye (e.g. SYBR™ Green) or a fluorescently labelled reporter probe, which allows the amplification of the target sequence to be monitored in real-time by the qPCR platform. Quantification is measured against a simultaneously ran

standard curve, based on samples of a known concentration of reference DNA (Herder et al. 2014). Probe-based qPCR increases both detection specificity and sensitivity, as the use of a probe, in combination with forward and reverse primers, ensures that there are three sequences to check against the target template DNA (Herder *et al.*, 2014). However, it is limited to the detection of only one or a few target organisms at a time (Sanger et al. 1977).

The alternative to traditional DNA Sanger sequencing (Taberlet et al. 2012) and eDNA barcoding (which can only sequence specimens individually) for species detection, is eDNA metabarcoding (the prefix 'meta' referring to the collection of barcode genes across the taxonomical spectrum of the samples). This multi-species approach simultaneously identifies multiple taxa from an environmental sample without the need for 'a priori' knowledge of the species likely to be present (Valentini et al. 2016; Yamamoto et al. 2017). Metabarcoding offers a tremendously enhanced capability in biodiversity studies because it has the potential to characterise the full community of species present in a set of complex environmental samples (R. P. Kelly et al. 2017; Leray & Knowlton 2015; Miya et al. 2015; Yamamoto et al. 2017). Metabarcoding employs high-throughput sequencing, while using more generalised PCR primers in order to mass-amplify a taxonomically informative marker gene and can thus offer a comprehensive view of an ecosystem. This method has the potential to reveal hundreds or thousands of taxa (and potentially their abundances) from a single environmental sample.

1.5 Applications of eDNA techniques in shark biology studies

One quarter of all Chondrichthyans (sharks, skates, rays, and chimeras) are currently considered threatened (Dulvy et al. 2014). Coastal and continental shelf-dwelling rays and sharks, such as sawfishes and angel sharks, are particularly at risk from overexploitation and other anthropogenic threats (Dulvy et al. 2016). Traditional survey methods have proven to be useful for determining the presence and distribution of a range of shark and ray species (Vaudo & Heithaus 2012; Kajiura & Tellman 2016; Kessel et al. 2016; Guttridge et al. 2017; Hansell et al. 2017). However, confirming the presence of a target species relies on locating and/or catching the animals, which can prove challenging and time-consuming for many species due to their rarity, cryptic habits, ecological specialisation and potential occurrence in remote and difficult to access locations (Barnes & Turner 2016). With over half of Chondrichthyan species considered data deficient, there is a clear urgency to rapidly increase the knowledge of these species' life histories and current distributional ranges to further

conservation and management efforts. Environmental DNA may be the "game changing" genetic technique for the study of sharks and their relatives, not only allowing for the time- and cost-effective gathering of crucial species' occurrence and distribution information, but also providing much needed ecosystem-wide species composition and population level data.

1.5.1 **Occurrence and distribution of rare and endangered species**

The application of eDNA techniques for the detection of sharks and their relatives has only been described recently. The first study dedicated specifically to the detection of an elasmobranch species, successfully detected the critically endangered largetooth sawfish (*Pristis pristis*) in freshwater habitats in northern Australia in locations with both known (based on gillnet surveys and traditional ecological knowledge from local Indigenous ranger groups) and unknown sawfish presence (Simpfendorfer et al. 2016). The versatility of the use of eDNA techniques in elasmobranch species detection has further been demonstrated with the eDNA barcoding approach being successfully applied in two widely different marine habitats: within a coastal embayment for the detection of the endangered Maugean skate (*Zearaja maugeana*) (Weltz et al. 2017), and above the summits of oceanic seamounts for the Chilean devil ray (*Mobula tarapacana*) (Gargan et al. 2017). Positive eDNA detections in water samples, identifying these endangered and critically endangered rays, respectively, highlights the value of the method for rare elasmobranch species. Moreover, the detection of oceanic and highly migratory species, such as devil rays, emphasises that species that are otherwise difficult and rare to encounter can be surveyed expeditiously with eDNA in open water environments.

Although still in its infancy, successful applications of eDNA barcoding for the detection of rare and endangered elasmobranchs in both marine and freshwater environments highlight the potential of this technique in informing conservation and management efforts. See Table 1.1 for a summary of shark related eDNA studies.

Table 1.1 Summary of eDNA studies to date focusing on sharks and their relatives. Details provided include climate zone and habitat, geographical location of study, genetic assay used, species of interest, collection and extraction methods, target gene and detection success.

Reference	Title	Climate Zone and Habitat	Geographical location	Study objective (Genetic Assay)	Species of interest	Collection method	Extraction	Target gene	Successful detection (no. positive detections/site)
Gargan *et al.* (2017)	Development of a sensitive detection method to survey pelagic biodiversity using eDNA and quantitative PCR: a case study of devil ray at seamounts	Temperate Seamounts	Azores	Species-specific (qPCR)	Chilean devil ray (*Mobula tarapacana*)	3 L collected per site, stored in -20 °C before vacuum filtration (0.45 µm nylon filter) and storage in 100% ethanol	DNeasy Blood & Tissue Kit (Qiagen)	COI	5/15
Weltz *et al.* (2017)	Application of environmental DNA to detect an endangered marine skate species in the wild	Temperate Harbour	Macquarie Harbour, Tasmania, Australia	Species-specific (qPCR)	Maugean skate (*Zearaja maugeana*)	2 × 10 L replicates stored on ice before vacuum filtration (0.45 µm cellulose nitrate filter) and storage in -18 °C	PowerWater DNA Isolation Kit	NADH 4	4/4
Bakker *et al.* (2017)	Environmental DNA reveals tropical shark diversity and abundance in contrasting levels of anthropogenic impact	Tropical Marine	Caribbean and New Caledonia. Multiple locations	Elasmobranch specific/ metabarcoding (Illumina amplicon sequencing)	All elasmobranchs	4 L samples, vacuum filtration (0.45 µm mixed cellulose esters (MCE) filters). Storage in silica beads at -20°C	Mo-Bio PowerSoil DNA Isolation Kit	COI	50/76
Boussarie *et al.* (2017)	Environmental DNA illuminates the dark diversity of sharks	Tropical Marine	New Caledonia	Elasmobranch specific/ metabarcoding (Illumina amplicon sequencing)	All elasmobranchs	4 L samples, vacuum filtration (0.45 µm mixed cellulose esters (MCE) filters). Storage in silica beads at -20°C	Mo-Bio PowerSoil DNA Isolation Kit	COI	21/22
Sigsgaard *et al.* (2016)	Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA	Tropical Eutrophic, high-saline Gulf	Al Shaheen, Central Arabian Gulf, off Qatar	Species-specific (qPCR and NGS)	Whale shark (*Rhincodon typus*)	3 × 500 mL replicates per site, stored on ice or filtered immediately using Sterivex-GP filters (0.22 µm) and storage in -18 °C	DNeasy Blood and Tissue Kit	D-loop	13/17
Simpfendorfer *et al.* (2016)	Environmental DNA detects Critically Endangered largetooth sawfish in the wild	Tropical Floodplain waterholes and main-channel river	Daly River, Northern Territory, Australia	Species-specific (PCR)	Largetooth sawfish (*Pristis pristis*)	5 × 2 L replicates per site, stored on ice prior to vacuum filtration (20 µm nylon filter) and storage in ethanol or -20 °C	ISOLATE II Genomic DNA Kit (Bioline)	COI	7/11
Thomsen *et al.* (2016)	Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes	Subarctic Continental slope	Southwest Greenland	Fish specific (Illumina amplicon sequencing)	Marine fish	1.5 L samples, vacuum filtration (0.45 µm nylon filters). storage at -20 °C	DNeasy Blood and Tissue Kit	12S rRNA	21/21
Miya *et al.* (2015)	MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species	Sub-tropical Aquaria	Okinawa, Japan	Metabarcoding (NGS)	Marine fishes	10 L replicates sub-sampled and vacuum filtered (0.7 µm glass fibre) and stored in aluminium foil in -20 °C	DNeasy Blood and Tissue Kit	12S rRNA	17/17 species present in aquaria
Kelly *et al.* (2014)	Using environmental DNA to census marine fishes in a large mesocosm	Sub-tropical Aquaria	Monterey Bay, California, US	Metabarcoding (NGS)	Marine fishes	20 L replicates sub-sampled and vacuum filtered (20 µm Durapore) and stored at -80 °C	DNeasy Blood and Tissue Kit	12S rRNA	No detection

1.5.2 **Species composition with eDNA metabarcoding**

Environmental DNA metabarcoding has the potential to simultaneously identify several taxa from environmental samples (Taberlet et al. 2012). Shark species inventories and assessment of geographical distributions based on eDNA metabarcoding could be an important tool for rapid environmental monitoring and hence influence conservation management and policy decisions. However, while three species of elasmobranch have previously been detected in a large-scale marine eDNA study, using a primer set designed for teleosts (bony fish) (Thomsen *et al*. 2016), other studies have encountered challenges concerning shark-specific detection, when applying this multi-specific approach in an aquarium-based setting (Kelly, J. A. Port, et al. 2014; Miya et al. 2015).

The main caveat in using eDNA metabarcoding for the assessment of shark diversity is that sharks are naturally rare compared to most other taxa. Figure 1.1 shows the difference in read abundances between elasmobranchs and teleosts recovered from two marine samples (from an area with relatively high shark abundances) and a sample taken from an aquarium tank, indicating the naturally low abundance of elasmobranch eDNA compared to teleost eDNA. Consequently, when individuals are present in the sampling area, eDNA released by sharks will generally only constitute a very small portion of all the eDNA present in a water sample. This highlights the importance of designing and optimising protocols specifically geared towards the detection of sharks. This includes sampling relatively large volumes of water (generally >3 litres) (Figure 1.2 A) per sample and using primers that specifically target sharks while excluding other, non-target, taxa.

More recently, we have employed eDNA metabarcoding of natural seawater samples to specifically infer shark presence, diversity and relative abundance in both Atlantic and Pacific tropical ecosystems (Bakker et al. 2017). By using a primer set targeting a 127 bp stretch of the mitochondrial COI region (Fields et al. 2015), twenty-one different shark species were detected, whose geographical patterns of diversity and abundance coincided with geographical differences in levels of anthropogenic pressure and conservation effort, in two independent tropical marine systems. Even though issues relating to the taxonomic assignment of closely related species still need to be resolved, this study demonstrates the potential of the eDNA metabarcoding approach for the detection and monitoring of shark communities.

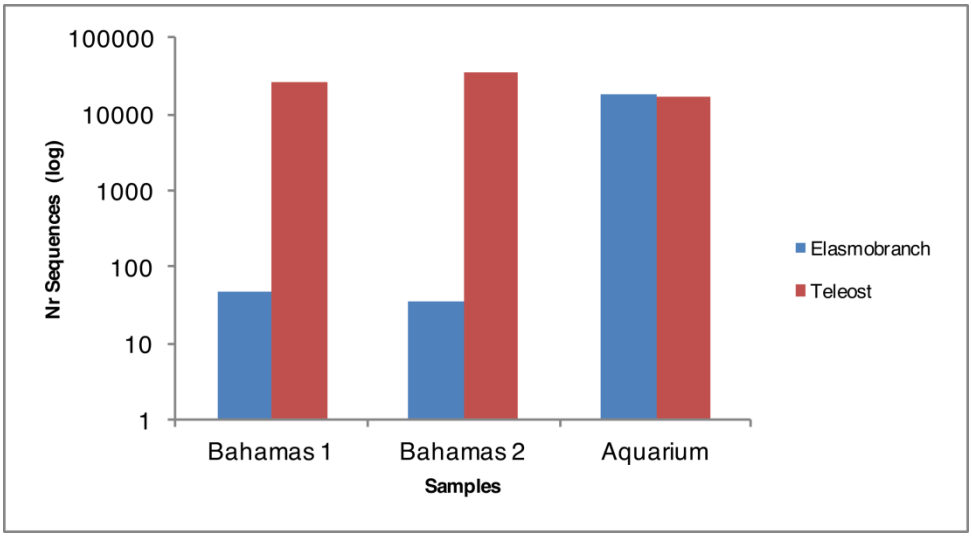


Figure 1.1 The differences in read abundances (on a logarithmic scale) between shark eDNA recovered from natural marine samples and an aquarium sample. Fish specific primers, targeting the control region (CR), were used for eDNA amplification.

1.5.3 Population genetics – From species detection to population analysis

An additional potential for the use of eDNA, if it stores sufficient population-specific information within the molecular markers used (e.g., mitochondrial haplotypes), lies in the area of population genetics, with applications for conservation genetics and phylogeography (Bohmann et al. 2014). To date, only one study applying eDNA to infer population characteristics for shark species, has been published (Sigsgaard et al. 2016). Here, samples were collected from areas in the Arabian Gulf, where whale sharks (*Rhincodon typus*) are known to occur. Mitochondrial DNA control region sequences obtained from eDNA samples were compared to sequences from tissue samples collected from the same locality. DNA mutation rate was calculated and female effective population size (N_f) inferred. Subsequently, it proved possible to infer the likely N_f for the entire Indo-Pacific Ocean, with comparable estimates obtained from eDNA and tissue extraction sequences. Moreover, by using eDNA analysis, this study revealed that the whale shark populations in the Indo-Pacific are genetically distinct from those populations occurring in the Atlantic Ocean. Thus, for the first time demonstrating that eDNA methods are capable of using the genetic variation in the DNA fragments isolated from water samples to estimate population sizes, as well as identifying relatedness between different populations of the same species.

1.6 eDNA Methods

Effective and accurate detection of organisms in aquatic ecosystems using eDNA is dependent on the development of an appropriate sampling design. There is no single eDNA sampling method that fits all target species and environments (Barnes & Turner 2016; de Souza et al. 2016), and conducting a pilot study is important before initiating a full study (Furlan et al. 2016; Goldberg et al. 2016; Kelly 2016). The method of water sample collection is the same for both species-specific and population-level investigations, but there are differences in field sampling design and downstream genetic processing and analyses. Overall, it is important to understand the characteristics of eDNA in the context of local environmental conditions, including the influence of biotic and abiotic factors on DNA degradation and dispersal; and factors related to the target species/community, including life history, demographic patterns and ecology. These factors can result in variation in detection sensitivity. Currently, the recommended protocol for each new application should assess detection probabilities for the target species given the proposed field and laboratory protocols (Goldberg et al. 2016). Preliminary laboratory and aquarium eDNA assays can be applied to test and confirm the sensitivity and specificity of the methodology and, where possible, controlled tank-based experiments conducted to further understand eDNA shedding, degradation, and distribution rates (Turner, Barnes, et al. 2014; Weltz et al. 2017).

Environmental DNA detection methods are perceived to be highly sensitive, but, logically, they are largely contingent on the probability of detecting eDNA where and when it is present in the environment (Ficetola *et al.*, 2008; Dejean *et al.*, 2012; Goldberg *et al.*, 2013). For sampling approaches that target a single species, estimating the sensitivity (or target species' detection probability) of the assay is crucial for accurately and confidently interpreting results, as it delineates the chances of detection failure (Amberg et al. 2015; Furlan et al. 2016). Detection failure; false positives (incorrect positive detection when the target species is absent) and false negatives (failing to detect the target species when it is present), potentially confound conclusions about species presence/absence and can misinform management. Therefore, the risk of such should be minimised through stringent execution of field and laboratory procedures.

The field sampling strategy for species-specific eDNA detection should consider the life history, behaviour and environment of the target species. Sharks and their relatives have diverse life history traits and occur in a vast array of marine, estuarine, and some freshwater

systems. Differences in habitat use will influence eDNA concentration and dispersion, and impact the likelihood of recovering target DNA from sample locations. Understanding the fine-scale patterns of occurrence and behaviour, such as movements and habitat use driven by ontogeny, predator avoidance, environmental tolerances, seasonal change, or fidelity, may allow enhanced detection ability. However, for many species this depth of information is lacking, and therefore data from similar species may render a useful tool to frame the development of an appropriate field sampling strategy. Moreover, where baseline information on patterns of occurrence and distribution does not exist, eDNA methods may be utilised as an exploratory tool to reveal this information.

When initially assessing field sampling strategy effectiveness or when targeting presence/absence information data at one point in time, sampling should occur during times and in locations a species is expected to be present (de Souza et al. 2016). For example, by utilizing existing knowledge on occurrence patterns oceanic, often solitary, deep-swimming elasmobranchs, such as devil rays, can be positively identified in oceanic basins, despite the dynamic and turbulent nature of ocean currents and wave action (Gargan et al. 2017). Contrarily, studies of resident species should have to account for variations in activity and behaviour in response to seasonally-variable factors such as temperature or precipitation, which, in turn, may influence eDNA abundance and persistence, and thus the probability of detection. Seasonal variations for example, may influence timing of reproduction of certain elasmobranch species, which is likely to increase the detectability of eDNA due to the release of reproductive material such as sperm, but also neonates (de Souza et al. 2016; Laramie et al. 2015; Spear et al. 2015). One might reasonably expect that discrete habitats used as pupping or nursery grounds by coastal elasmobranchs would contain higher proportions of eDNA as a result of reproductive behaviours. Likewise, increased activity during tidal- or diurnal-driven movements or feeding behaviour may also increase eDNA shedding rate. Species-level differences in habitat use and behaviour may dictate spatial and temporal considerations for eDNA detection.

1.6.1 **Water collection, filtration, preservation & extraction**

Capturing eDNA from an aquatic environment is the crucial first step in the eDNA workflow. Environmental DNA starts to decay immediately after shedding and continues to do so after sample collection (Turner, Barnes, et al. 2014; Dejean et al. 2012; Pilliod et al. 2014;

Sassoubre et al. 2016; Yamanaka et al. 2016). A recent study on eDNA recovery rates following various combinations of eDNA capture, preservation and extraction methods has indicated that DNA yield (copy number) from stream water samples, prior to filtration, significantly decreases when stored at either room temperature (20˚C), refrigerated (4˚C) or frozen (-20˚C) from day 1to day 2, regardless of storage temperature (Hinlo et al. 2017). Moreover, in a different study, using decay modelling of Maugean skate (*Zearaja maugeana*) eDNA, Weltz *et al*. (2017) showed that the eDNA concentration in some water samples had fallen below the detection limit of the assay within 4 hours of sampling. Therefore, samples should be filtered and extracted, and the eDNA extracts preserved using prescribed protocols, as soon as possible after water collection. Precipitation and filtration are the most commonly used methods to recover eDNA from water samples. Other methods include preservation of small volumes of water followed by concentrating the DNA by centrifugation (Klymus et al. 2015).

Generally, precipitation involves the collection of small volumes of water (e.g. 15 mL) (Ficetola *et al.*, 2008; Eichmiller, Miller and Sorensen, 2016) that are immediately preserved in-field with the addition of sodium acetate and absolute ethanol (salt and ethanol precipitate nucleic acids from water (Maniatis *et al*., 1982), prior to storage at -20C. The precipitation method requires few collection tools (i.e. precipitation solution and collection vials), thus the relative ease of this method is a major benefit for users. Where the processing of larger volumes of water is required, it is advisable to increase the number of biological replicates or, alternatively, use the filtration method. Filtration is more advantageous when dealing with larger bodies of water such as rivers, estuaries or marine environments (Hinlo et al. 2017; Turner, Miller, et al. 2014).

Filtration requires the passage of water through a membrane that captures the eDNA, and generally allows the processing of larger volumes of water (typically 1-10 l). Filtration can be carried out on-site with a portable filtration system (Figure 1.2 A), or water samples can be stored on ice and transported to a laboratory (or equivalent processing facility) for filtration. If not performed in the field, filtration should be undertaken as soon as possible (i.e. within 24 hours) to ensure optimal eDNA recovery (Hinlo et al. 2017; Weltz et al. 2017).

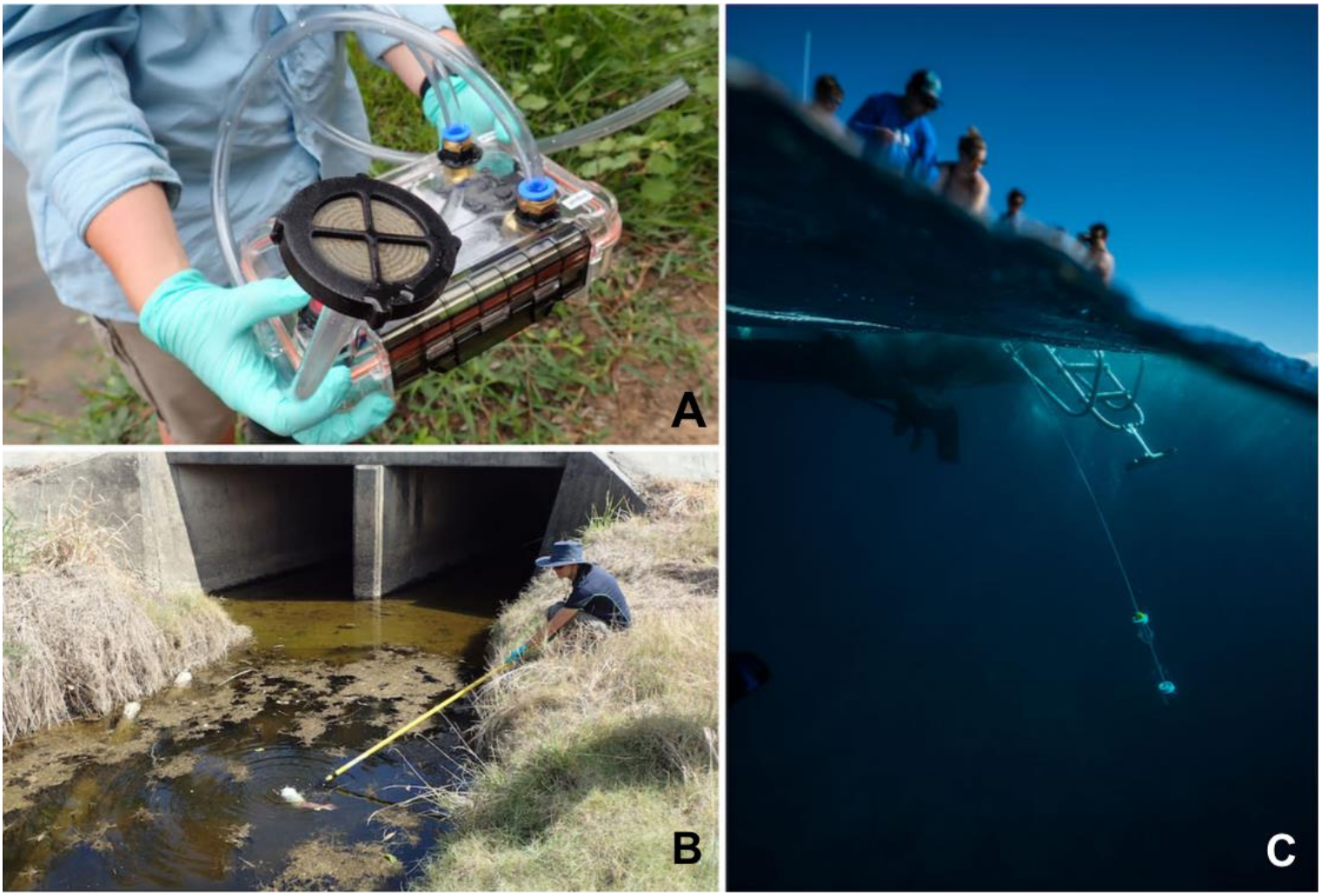


Figure 1.2 Field equipment used in eDNA studies: (**A**) Portable eDNA filtration pump (Robson pump) that can be used to quickly filter samples on-site. Photograph credit: Madie Cooper; (**B**) Extendable pole used in difficult to reach areas or to decrease the risk to the sampler of dangerous wildlife (e.g. crocodilians); (**C**) Collection of ocean water with a Kemmerer type water sampler. Photograph credit: Diego Camejo

Depending on field conditions, cold storage of filters wrapped in aluminium foil or contained in sterile microcentrifuge tubes may not be practical. However, this is commonly employed in laboratory-based settings or where field locations are close to the laboratory. When field conditions preclude the use of refrigeration, ethanol is the most commonly used alternative for filter preservation and storage. Other ambient temperature buffers, such as Longmire's solution and cetyl trimethyl ammonium bromide (CTAB), have been successfully used to preserve eDNA contained on filters (Spens et al. 2017; Williams et al. 2016; Renshaw et al. 2015), but require preparation using several ingredients, and the latter is a toxic substance. Longmire's buffer can also be used to preserve small volumes of unfiltered water at ambient temperature for up to 56 days prior to DNA extraction (Williams et al. 2016). An alternative method for eDNA preservation is adding silica beads to the vessel containing the filter, the beads function as a desiccator, drying out the filter and preventing the DNA from degrading (Bakker et al. 2017). Long-term eDNA recovery rates from ethanol and other preservatives is currently unclear and further research is needed, however we have successfully recovered eDNA from filters, desiccated with silica beads, after >1 year of storage at -20 °C.

The type of filter membrane used for the separation of eDNA from the environmental samples also varies. Glass fibre, nylon, cellulose nitrate, polycarbonate, polyethersulfone, and cellulose acetate filters have previously been used (Deiner et al. 2015; Goldberg et al. 2016; Renshaw et al. 2015). The inherent properties of the filter material [i.e. depth filters: particles retained on the surface and within the filter matrix, versus surface filters: particles trapped on filter's surface (Hinlo et al. 2017), affect the binding affinity of eDNA and, as such, eDNA recovery rates differ, depending on the type of filter used (Liang & Keeley 2013).

Filter pore size is an additional factor to consider when choosing filters for filtration-based eDNA recovery. Intuitively, larger sample volumes will increase eDNA capture success. However, there is a trade-off between sample size, pore size, and eDNA particle retention; a smaller pore size captures more eDNA particles but limits sample volume and speed. Conversely, a larger filter pore size allows for a faster flow rate and larger sample volume at a faster filtering rate, but in turn may reduce the amount of eDNA particles captured on the filter. Hence, there are two important considerations to take into account when choosing the correct filter pore size: size distribution of eDNA particles, and water turbidity at the sampling location. Knowledge of the size distribution of various intra- or

extracellular eDNA particles will assist in informing on the trade-off between filter pore size and sample size. Turner *et al* (2014a) observed size fractions of common carp (*Cyprinus carpio*) eDNA and concluded that the largest amount of total eDNA recovered was within the 1-10 µm size fraction. Comparable studies for sharks and their relatives do not currently exist, and it is unclear whether the aforementioned findings are representative of general size distributions for all eDNA or are taxa- or environment-specific. With this in mind, small pore sizes should be used where possible to ensure highest possible eDNA capture rate. For example, filter pore sizes ranging from 0.45 to 3 µm are most commonly used in studies undertaken in less turbid water (Gargan et al. 2017; O'Donnell et al. 2017; Sigsgaard et al. 2016; Weltz et al. 2017). For more turbid water, however, even 3 to 5 µm filters quickly become clogged with suspended particulate matter, necessitating the use of larger pore sizes of up to 20 µm to minimise clogging and maintain an efficient filtration rate (Robson et al. 2016; Simpfendorfer et al. 2016). If filter clogging is a frequent occurrence, multiple filters may be used and eDNA extracts pooled for sample replicates.

Multiple Different eDNA extraction methods can be applied to isolate eDNA captured by filtration or precipitation, but also to remove compounds that can inhibit downstream enzymatic reactions such as PCR (Eichmiller et al. 2016). Inhibitors may range from cellular components to materials in the water, such as humic substances (Wilson 1997), that are captured together with the eDNA. Both capture methods may be followed by either Phenol-Chloroform-Isoamyl alcohol (PCI) DNA extraction or extraction using a commercial DNA extraction kit (Deiner et al. 2015). The DNA extraction kits such as the commonly used Qiagen DNeasy Blood & Tissue Kit and Qiagen's DNeasy Power Water (and Power Soil/Power Max) DNA isolation Kits, are convenient and simple to use but are more expensive compared to PCI extraction. While PCI extraction in turn requires careful preparation and handling of toxic chemicals. Several studies have found that PCI extraction yields more eDNA compared to commercial DNA extraction kits (Deiner et al. 2015; Renshaw et al. 2015; Turner et al. 2015). While another study observed more PCR inhibition in DNeasy extracted samples compared to Power Water extracted samples (Eichmiller et al. 2016), which is likely a result of different additives to alleviate the effects of PCR inhibitors. Hence, high eDNA yield does not necessarily accompany increased species detection but is rather dependent on a multitude of factors. Likewise, Deiner *et al*. (2015) have demonstrated that different combinations of eDNA capture and extraction protocols result in different detection rates of biodiversity.

Environmental characteristics (e.g. water chemistry and temperature), target species, capture method, filter material and pore size, storage, and DNA extraction method interact to produce final detection rates (Deiner et al. 2015; Eichmiller et al. 2016; Goldberg et al. 2016; Renshaw et al. 2015), and there is no one extraction method that is equally beneficial to all taxa and/or ecosystems for the maximisation of eDNA recovery and target species detection. Thus, it is recommended that different combinations of storage, preservation, filter type and extraction methods are tested and optimised, depending on the research objectives, preference, ease of use, and availability of resources. Finally, detailed information about the field, laboratory and bioinformatic procedures used in eDNA studies should be reported to enhance the development of the field by increasing communication about techniques and quality control.

1.6.2 **Selecting gene regions for target organisms or groups**

Methodologically, eDNA detection requires the development of genetic markers specific to the target taxon or taxa. Targeted eDNA fragments may be detected using different molecular methods including Sanger sequencing, qPCR and (meta)barcoding. When aiming to detect a single species, primers should be specific to the target species, while incorporating as many differences as possible to other sequences of related organisms (Ficetola *et al.*, 2008). Insufficient primer specificity can lead to over or under estimation of species presence and, especially when taxa closely related to the target species are present, cross-amplification or interference of amplification can lead to the generation of false positive and negative errors (Wilcox et al. 2013).

Target loci are typically within the mitochondrial genome because of its greater biological abundance and higher level of coverage in genetic databases. However, selecting the correct gene region for a targeted eDNA barcoding approach will ultimately depend on how much intra- and interspecies variability is found for the species of interest at a particular gene. Environmental DNA barcoding studies to date have designed species-specific assays within a wide range of genes including cytochrome B (Spear et al. 2015; Wilcox et al. 2013; Hunter et al. 2015), cytochrome oxidase I (COI) (Gargan et al. 2017; Nathan et al. 2014; Simpfendorfer et al. 2016; Brandl et al. 2015), nicotine adenine dinucleotide dehydrogenase

subunit 4 (NADH4) (Hunter et al. 2015; Weltz et al. 2017), 16S (Robson et al. 2016), and 12S (Furlan and Gleeson 2016a; Secondi *et al.*, 2016).

In studies where a large number of species co-occur, some of which may be closely related, finding a suitable gene to design a species-specific or even genus-specific primer assay may be challenging. This may also be true for sharks and their relatives. Sharks, and most likely also rays, appear to have slow mutation rates in mtDNA compared to other vertebrates (i.e. mammals and teleost fish) that lead to lower genetic variation (Martin 1995; Martin 1999; Martin et al. 1992). For example, mitogenomic sequencing in the critically endangered speartooth shark (*Glyphis glyphis*) has revealed one of the lowest known levels of genetic diversity (Feutry *et al*. 2014). Increasingly, primer assays are being designed using whole mitogenome sequencing to find suitable gene regions (Hunter et al. 2015), as this increases the chances of finding suitably variable gene regions, potentially in less commonly used alternative regions that exhibit useful polymorphisms.

In contrast, when choosing a suitable genetic marker for metabarcoding, a genomic region with sufficient sequence variability must be targeted in order to be able to distinguish closely-related species. It must be flanked by conserved regions, which act as primer attachment sites. Moreover, a region with many copies per cell is preferable, as this natural abundance of DNA sequences will facilitate amplification. Hence, organelle genomes, such as mitochondrial or chloroplast DNA, or ribosomal RNA clusters, are usually preferred targets (Wangensteen et al. 2017; Wilcox et al. 2013). For eDNA applications, the target fragment length must be relatively short, as eDNA released in the environment rapidly degrades into small fragments, thus the chances of amplifying the full length of the marker from eDNA is inversely proportional to the length of the chosen marker (Wangensteen et al. 2017). For eDNA metabarcoding, additional considerations apply, as the most popular method for eDNA high-throughput sequencing, the Illumina platform, currently has a maximum effective read length of around 500 bp; however, in order to keep sequencing error rates low, smaller fragments are preferred. Therefore, the ideal length for an eDNA metabarcoding marker should not exceed 350 bp.

The universality or specificity of the primer set is dependent on the breadth of the taxonomic scale of interest. For example, primer sets for the elucidation of elasmobranch (Bakker et al. 2017), teleost (Miya et al. 2015) or arthropod (Zeale et al. 2011) diversity, can be used. Conversely, targeting whole eukaryotic community diversity will require a primer set that is as universal as possible in order to be able to attach to the marker flanking sequences in most taxonomic groups, so that all these groups will be adequately amplified by

PCR. As of yet, there is no ideal universal metabarcode that is able to amplify the full taxonomic range of a community, for highly variable markers such as COI (Coissac et al. 2012; Deagle et al. 2014; Riaz et al. 2011). Thus, truly universal primers have usually been restricted to markers with more conserved regions such as 18S (Guardiola et al. 2015). However, the development of primers including deoxyinosine (a nucleotide which complements any of the four natural bases) in the fully degenerated sites of the sequence may improve the universality of COI primer sets (Wangensteen, Palacín, Guardiola 2017)

The use of COI as a metabarcoding marker has previously been criticised, arguing that high rates of sequence variability impair the design of truly universal primers and hampers bioinformatics analysis. Instead, mitochondrial rRNA genes have been recommended for animal identification because they have a similar taxonomic resolution as the COI marker and they present conserved regions that flank variable regions, which allows the design of primers with high resolution power for the target taxonomic group (Deagle et al. 2014). However, it may be argued that COI presents two major advantages over other potential markers. First, the steadily growing international effort, headed by the Consortium for the Barcode of Life (CBOL), to develop a public DNA barcoding database with curated taxonomy, greatly facilitates taxonomic assignment. The BOLD database (http://www.boldsystems.org/) (Hebert, Ratnasingham, et al. 2003; Ratnasingham & Hebert 2007) currently includes >4 million sequences belonging to over 500,000 species, curated and identified by expert taxonomists. Secondly, the high mutation rate of COI ensures unequivocal identification at the species level, which is crucial for studies aimed at detecting rare or invasive species, such as may be the case for sharks. Whereas the highly conserved sequences of other markers, such as 18S, make it often impossible to distinguish at the species or genus levels.

1.7 The challenges of eDNA studies

1.7.1 Contamination

One of the main challenges associated with the use of eDNA, is dealing with false positive and false negative detections (Darling & Mahon 2011). Due to the high sensitivity of eDNA methods, the most serious stumbling block is the risk of contamination (Goldberg et al. 2016;

Thomsen & Willerslev 2015), and hence the possibility of introducing false positive results. Contamination of samples may occur anywhere from preparing sampling equipment and collecting the samples in the field (target DNA being carried unintentionally from one locality to another), to every subsequent step of sample preparation, DNA extraction and analysis in the laboratory. Due to the frequent use of PCR, generating billions of DNA copies, contamination occurring in the laboratory can potentially have serious implications for the resulting data set, with important downstream repercussions on conservation and management decisions resulting from these results. Thus, precautions must be taken at all stages by putting strict procedures in place both in the field (establishing clean and consistent field collection protocols), and in the laboratory (implementing strict, clean lab protocols) in order to prevent the occurrence of contamination. This includes the use of disposable gloves and the disinfection/bleaching of sampling devices and all laboratory equipment. Additionally, filtration, DNA extraction and PCR procedures, as well as pre- and post-PCR procedures, have to be separated physically to limit the risk of contamination (Goldberg et al. 2016; Wilson et al. 2015). Moreover, to monitor potential contamination (i.e. to identify the source of contamination when it occurs), the inclusion of field blanks (clean water sampled using the same protocol and equipment, preserved and processed in exactly the same way as the actual field samples), DNA extraction blanks and PCR blanks, is essential (De Barba et al. 2014).

1.7.2 **eDNA shedding rates**

The availability of detectable eDNA in environmental samples is reliant on the underlying premise that all organisms shed genetic material. Earlier studies on terrestrial vertebrate eDNA detection in aquatic environments imply the most probable origin of eDNA as faecal material (Martellini et al. 2005). While this may be remain true for a wide range of taxa, the origin of eDNA from aquatic organisms is also linked to species-specific physiological characteristics such as skin properties (e.g. slimy coatings on amphibians (Ficetola *et al.*, 2008) and fish (Jerde et al. 2011), metabolic rates (Klymus et al. 2017), reproductive mode and timing (Spear et al. 2015; Bylemans et al. 2017), feeding rates (Sassoubre et al. 2016), and environmental tolerance (Lacoursiere-Roussel et al. 2016; Robson et al. 2016). The composition of eDNA containing genetic material from these origins remains relatively unclear and particularly hard to study, however, many complex factors influence eDNA

shedding rate and, as such, interpretation of eDNA detection results benefits from a complete understanding of the ecology of eDNA.

Overall, in marine and freshwater organisms, it is largely understood that eDNA shedding rates are foremost positively related to individual and/or population biomass (Pilliod et al. 2014; Stoeckle et al. 2017; Thomsen et al. 2016; Weltz et al. 2017). It is this correlation that underpins the use of eDNA concentration in water as a proxy measure of biomass of the focal organism/s, which has been applied to both species-specific (Sigsgaard et al. 2016; Doyle et al. 2017) and population-level questions (Kelly 2016; Leray & Knowlton 2015; Miya et al. 2015; Yamamoto et al. 2017). In tank-based experiments, eDNA release rates demonstrated linear positive relationships with biomass (Klymus et al. 2015; Sassoubre et al. 2016). However, it is likely that this relationship is more complex, as highly variable eDNA production rates among individuals, unrelated to biomass, have been observed, suggesting that this variation may be attributable to animal physiology (Buxton et al. 2017; Maruyama et al. 2014; Pilliod et al. 2014; Wilcox et al. 2016). For juvenile fishes, it was concluded that ontogenetic factors, such as differences in behaviour and metabolism, increased eDNA shedding rates per body weight compared to adult conspecifics (Klymus et al. 2015; Maruyama et al. 2014). Moreover, it has been suggested that stress and feeding behaviour can influence eDNA shedding rates (Sassoubre et al. 2016), while these behaviours are intertwined with the physiological tolerances of aquatic organisms (Lacoursiere-Roussel et al. 2016). Additionally, the eDNA contribution from different life stages may vary seasonally. For example, strong temporal increases in eDNA concentration have been observed during months associated with seasonal migration and breeding (Buxton et al. 2017; Doi, Uchii, et al. 2015; Spear et al. 2015; Fukumoto et al. 2015).

Seasonal migrations and patterns of occurrence associated with specific behaviours are dictated by metabolic function, which, in most sharks, is determined by water temperature. As water temperature increases, mobility and metabolic rate increases until the upper limit of physiological tolerance is reached. Some sharks also perform diel vertical migrations to conserve energy in deeper, cooler waters and search for prey when near the surface in warmer water (Sims et al. 2006). Increased water temperature and digestive function, coupled with movements associated with prey-seeking and feeding behaviour, leads to the increased excretion of metabolic waste and the release of epidermal cells containing genetic material. While some studies have shown no effect of temperature on accumulation or shedding of eDNA in fishes (Klymus et al. 2015; Takahara et al. 2012), a recent study on the tropical invasive fish species tilapia (*Oreochromis mossambicus*) showed increased

shedding rates of eDNA at 35°C, a temperature well within their known thermal tolerance (Robson et al. 2016). Moreover, estimates of fish biomass in aquarium samples were better reflected in warmer water as supported by higher eDNA concentration and shedding rates from fish in warm water compared to colder water (Lacoursiere-Roussel et al. 2016).

Other physiological attributes such as skin properties provide clues to possible origins of eDNA-bearing particles. High eDNA detection success rates have been observed for fish that produce slimy coatings (Jerde et al. 2011). Comparably, sharks and their relatives also produce epidermal mucus (Meyer & Seegers 2012; Tsutsui et al. 2009). While the mucus layer on the skin surface of demersal sharks and rays (e.g. angel sharks, *Squataina spp*.) is comparatively thicker than on pelagic, fast moving, predatory sharks, this does suggest that mucus-derived genetic material may contribute to a portion of all elasmobranch eDNA.

1.7.3 **eDNA degradation**

Environmental DNA possesses limited chemical stability (Lindahl 1993); once shed from an organism it begins to degrade into small fragments and becomes undetectable within hours to weeks (Dejean et al. 2011; Piaggio et al. 2014; Pilliod et al. 2014). Degradation is the primary mechanism limiting detection of species through eDNA. However, due to the short lifespan of eDNA it is thought to provide approximate real-time data on species presence in the environment. The persistence of eDNA for aquatic taxa has been estimated at 15 to 30 days for freshwater fishes (Dejean et al. 2011; Takahara et al. 2012) and hours to 7 days for marine fishes (Thomsen, Kielgast, Lars Lønsmann Iversen, et al. 2012), after which time eDNA concentrations drop below the detection limit. In specific reference to sharks and their relatives, it is accepted that eDNA exponentially decays in aquatic environments and becomes undetectable within hours to 5 days (Sigsgaard et al. 2016; Weltz et al. 2017). For example, in controlled degradation experiments, concentration of whale shark **(***Rhincodon typus*) eDNA dropped an order of magnitude within the first 48 hours, and was no longer detectable 8 days post-sampling (Thomsen et al. 2016), and Maugean skate (*Zearaja maugeana*) eDNA remained detectable for up to 5 days (Weltz et al. 2017).

Environmental conditions play an integral role in eDNA persistence and degradation (Barnes et al. 2014a). Understanding the interactions of environmental factors controlling degradation is essential to inferring the limits of temporal and spatial inference of eDNA detection results. Drivers of eDNA degradation are classified into three categories: (1) DNA

characteristics, including length, conformation, and association with membranous material (Taberlet et al. 2012; Willerslev et al. 2009), (2) abiotic environment, including temperature, pH, UV radiation, oxygen, and salinity (Barnes et al. 2014a; Pilliod et al. 2014; Strickler et al. 2015; Weltz et al. 2017), and (3) biotic environment, including exogenous enzymes and microbial activity (Dejean et al. 2011).

Fragments of DNA in the environment occur in different lengths, sequences, and conformations, which influence how eDNA binds to other particles and interacts with microbes in the environment (Lennon 2007; Ogram et al. 1988), which increases and decreases the rate of degradation, respectively. Binding to sediment particles can play a role in both eDNA concentration and preservation in surficial sediments (Turner et al. 2015), as does containment within cellular or organelle membranes, by providing protection from external degradative forces. Marine sediment eDNA concentrations have been shown to be three orders of magnitude higher than that of eDNA in seawater (Torti et al. 2015). Moreover, DNA has a stronger affinity for clay particles compared to sand or silt (Romanowski et al. 1992), and while sediments are not typically sampled in a presence/absence or contemporary occurrence survey because of the longevity of eDNA, consideration should be given to the potential resuspension of sediments in the water column and the subsequent increased probability of detecting this preserved eDNA.

Marine and freshwater tropical environments have high surface temperatures (sometimes above 30°C) and elevated UV radiation at sea level that may increase eDNA degradation rate and reduce its persistence in the water, decreasing the detection probability (Barnes et al. 2014a). Higher temperatures can denature DNA molecules, albeit when temperatures are >50°C, and indirectly increase microbial metabolism and exogenous nuclease activity (Fu et al. 2012; Kreader 1998). For example, Robson *et al.*, 2016 showed that high water temperatures of up to 35°C did not affect eDNA degradation rates of the invasive Mozambique tilapia (*Oreochromis mossambicus*). Ultraviolet radiation (UV), particularly UV-B, can directly damage DNA (but see (Andruszkiewicz et al. 2017) and has variable effects on exogenous nuclease production indirectly inhibiting eDNA persistence.

Acidic, hypersaline, or anoxic environments can influence eDNA stability and increase degradation rate. Deviation from neutral pH may reduce degradation rates, especially pertaining to the pH requirements of extracellular microbial enzymes that are considered to have a large impact on eDNA degradation (Sigsgaard et al. 2015). High salinity of samples may have negative downstream effects, such as inhibition of PCR, but this can be mediated by adding an ethanol wash step to the extraction process in order to remove

monovalent Na+ ions (Foote et al. 2012). The interaction of all biotic and abiotic factors combined is likely to have variable and synergistic effects on the mechanism of eDNA degradation in aquatic systems.

1.7.4 **eDNA capture rate**

Understanding the physical movement of eDNA in the environment is essential for correctly inferring presence of organisms in space and time and, hence, drawing robust conclusions within spatial and temporal boundaries (Barnes & Turner 2016). Environmental DNA represents a complex mixture of particles ranging in size and composition, which behave independently and move freely in aquatic environments. These particles are randomly and heterogeneously distributed in the water column as a result of spatial clumping (Furlan et al. 2016). The greater the degree of clumping and uneven dispersal of target DNA, the greater the likeliness of false negatives. Consequently, detection sensitivity for a given sampling protocol may vary temporally and spatially, between samples and from site to site, depending on the concentration and dispersion of target eDNA (Furlan et al. 2016; Weltz et al. 2017). Differences in eDNA detection sensitivity across space and time may also be the result of the differences in activity levels or other site-level factors that influence eDNA concentration, including the density or biomass of the target species. The use of hierarchical occupancy and detection sensitivity models, that take into account the specific survey methods used, both in the field and laboratory, may be applied to optimise capture and detection probabilities (Furlan et al. 2016; Schmidt et al. 2013).

1.7.5 **eDNA transport: lentic versus lotic systems**

Although the high sensitivity of eDNA assays in mesocosms and lentic systems (still waters) is well established (Thomsen, Kielgast, Lars L. Iversen, et al. 2012), studies in lotic systems (flowing waters) have more varied results, with potentially important management and conservation implications. This is typically illustrated by eDNA assays that show high detection rates (100%) when tested in ponds, but have much lower detection rates once used in the target species' natural lotic environment (54%) (Thomsen, Kielgast, Iversen, *et al.*, 2012).

Environmental DNA may occur at very low concentration in the aquatic environment and can be heterogeneously distributed. Thus, knowledge of how eDNA distribution is affected by water movement (e.g., currents, eddies, waves) and what additional interacting external drivers may affect its detectability; e.g., abiotic and biotic factors involved in eDNA persistence in the environment (Barnes et al. 2014a; Strickler et al. 2015; Jane et al. 2015) is crucial for successful detection. This is particularly the case for the detection of rare species, for which eDNA concentrations are likely to be at their lowest (Takahara et al. 2012), and the risk of false negative errors high.

Long distance transport of eDNA from hundreds of meters to several kilometres has been reported in river systems, and should always be taken into account in eDNA studies in lotic systems (Deiner & Altermatt 2014; Jane et al. 2015). Although it might be expected that eDNA may travel much larger distances in highly dynamic systems such as open oceans or flowing rivers compared to more stagnant systems such as ponds and lakes (Deiner & Altermatt 2014; Shogren et al. 2016), recent work on a dynamic marine coastline found evidence that eDNA transport was limited to the extent that eDNA metabarcoding methods were able to detect differences among vertebrate communities separated by less than 100 m (Port et al. 2016). Additionally, Gargan *et al.*, (2017) have detected the Chilean devil ray (*Zearaja maugeana*) using a targeted eDNA approach at 4 out of 5 remote seamounts that were sampled around the Azores, which was consistent with visual observation data. However, failure to detect target eDNA at a location where the species had been observed, highlights the influence of detection stochasticity and the need for further investigations into how eDNA transport and degradation affects species detection in open ocean environments.

1.7.6 **Freshwater vs. seawater**

For both freshwater and marine ecosystems, eDNA detection is correlated with the abundance of the target species and the rate at which DNA is released and degraded by biotic and abiotic factors (Thomsen, Kielgast, Lars Lønsmann Iversen, et al. 2012; Thomsen, Kielgast, Lars L. Iversen, et al. 2012). A considerable amount of aquatic eDNA research has been focused in freshwater systems (e.g. Ficetola *et al.*, 2008; Dejean *et al.*, 2011; Jerde *et al.*, 2011; Takahara *et al.*, 2012; Gustavson *et al.*, 2015; Laramie, Pilliod and Goldberg, 2015). Only more recently have eDNA studies been focused on species detection in seawater samples (e.g. Foote *et al.*, 2012; Thomsen, Kielgast, Lars Lønsmann Iversen, *et al.*, 2012;

Sigsgaard *et al.*, 2016; Thomsen *et al.*, 2016; Gargan *et al.*, 2017; Weltz *et al.*, 2017). Although a large proportion of these have been carried out in controlled environments, such as aquarium tanks (Foote et al. 2012; Kelly, J. A. Port, et al. 2014; Miya et al. 2015), more recently, successful eDNA studies involving natural marine environments have been reported. These studies have covered a variety of environments; coastal waters (Weltz et al. 2017; Yamamoto et al. 2017), open-ocean seamounts (Gargan et al. 2017), offshore oil fields (Sigsgaard et al. 2016), and continental slope depths (Thomsen et al. 2016).

The current lag in eDNA studies in marine ecosystems may stem from the perception that species detection from seawater samples may be more challenging compared to those from freshwater, inherent to the larger body of source water and, tidal and current action potentially rapidly diluting and dispersing the eDNA likely up to 100s of kilometres away (Thomsen, Kielgast, Lars Lønsmann Iversen, et al. 2012). But see (O'Donnell et al. 2017; Port et al. 2016). Environmental DNA is not only subject to transport, but also to degradation from exposure to various biotic and abiotic stressors (i.e., temperature, salinity, pH, UV-B, enzymes) (Goldberg et al. 2016). However, how these combined factors affect the potential of eDNA techniques to detect marine organisms in coastal or open ocean environments has seldom been investigated, and most marine studies in this field have only focused on determining eDNA shedding and decay rates in a handful of species (Andruszkiewicz et al. 2017; Sassoubre et al. 2016; Sigsgaard et al. 2016; Thomsen, Kielgast, Lars L. Iversen, et al. 2012). DNA degradation in seawater has previously been suggested to be substantially faster than in freshwater, with an empirical turnover rate as low as 10 hours (Dell'Anno & Corinaldesi 2004). Although abiotic and biotic stressors in the marine environment are likely to differ from those impacting freshwater systems (Thomsen & Willerslev 2015), it is not clear which of these are mostly responsible for the increased rate of eDNA degradation. A recent study investigating the impact of sunlight (UVB and combined UVA+UVB radiation) on the decay of Pacific chub mackerel (*Scomber japonicus*) eDNA in a marine water mesocosm, concluded that sunlight was not an important factor in promoting eDNA degradation, and suggested that factors other than sunlight, such as bacteria, grazers and enzymes are likely to have a more substantial impact (Andruszkiewicz et al. 2017). Moreover, another experimental study has suggested that the time it takes for Maugean skate (*Zearaja maugeana*) eDNA to degrade beyond its detection limits, was influenced by the dissolved oxygen (DO) concentration in the eDNA sample (Weltz et al. 2017).

Environmental DNA in freshwater systems has been shown to degrade beyond the threshold of detectability within a short time-frame (days to weeks) (Dejean et al. 2011;

Piaggio et al. 2014; Pilliod et al. 2014; Thomsen, Kielgast, Lars L. Iversen, et al. 2012), providing a "real-time" measure of species presence. Conversely, in the marine environment, eDNA may decay below the detection threshold in as little as 4 hours post-sampling (Weltz et al. 2017). Other studies have also indicated slightly slower rates of degradation, at a scale of days (Andruszkiewicz et al. 2017; Sigsgaard et al. 2016). As the rate of degradation of eDNA is inherently linked to both the starting concentration and the abiotic factors promoting degradation (e.g. UV, pH, etc.), it is possible that one of the challenges of eDNA studies in seawater is the inherently greater dilution of the eDNA signal.

Field and laboratory practices for the application of eDNA analysis to seawater may be modified to counter some of the aforementioned challenges. Larger volumes of water and a larger number of field replicates within a study area may be collected to counter for the greater water volume:biomass ratio of marine systems. Since eDNA concentrations are expected to be lower in the open ocean than in river systems, this is particularly pertinent when dealing with species that are likely to occur in low numbers or that are sparsely distributed. Additionally, caution should be exercised when using eDNA concentrations for the estimation of abundance of fish species in marine systems until further work is carried out, elucidating the persistence of eDNA under the influence of biological, environmental and physical processes, and how processes such as eDNA shedding, degradation, and transport can be integrated into reliable estimates of abundance.

1.7.7 **eDNA in tropical ecosystems**

Since the introduction of eDNA into mainstream environmental research, the majority of eDNA studies have been applied to temperate systems (Thomsen & Willerslev 2015). Comparatively little research has been applied to tropical aquatic systems, and the applicability and reliability of eDNA methods for effective species and community detection, and conservation management in these kind of environments is less clear.

The tropics present their own sets of challenges, with eDNA in marine and freshwater tropical environments exposed to more extreme conditions for longer and more frequent periods of time. Tropical aquatic systems have high surface water temperatures (sometimes >40°C), elevated UV radiation at sea level, and higher levels of microbial activity that may increase eDNA degradation rates and reduce their persistence in the water (Barnes et al. 2014a). Furthermore, seasonal precipitations (wet season) may lead not only to increased

turbidity due to high sedimentation and algae loads, but also to increased dilution due to high water flow rates (Figure 1.3A&B). The interaction of these factors, specific to tropical systems, is likely to significantly influence the detection of eDNA.

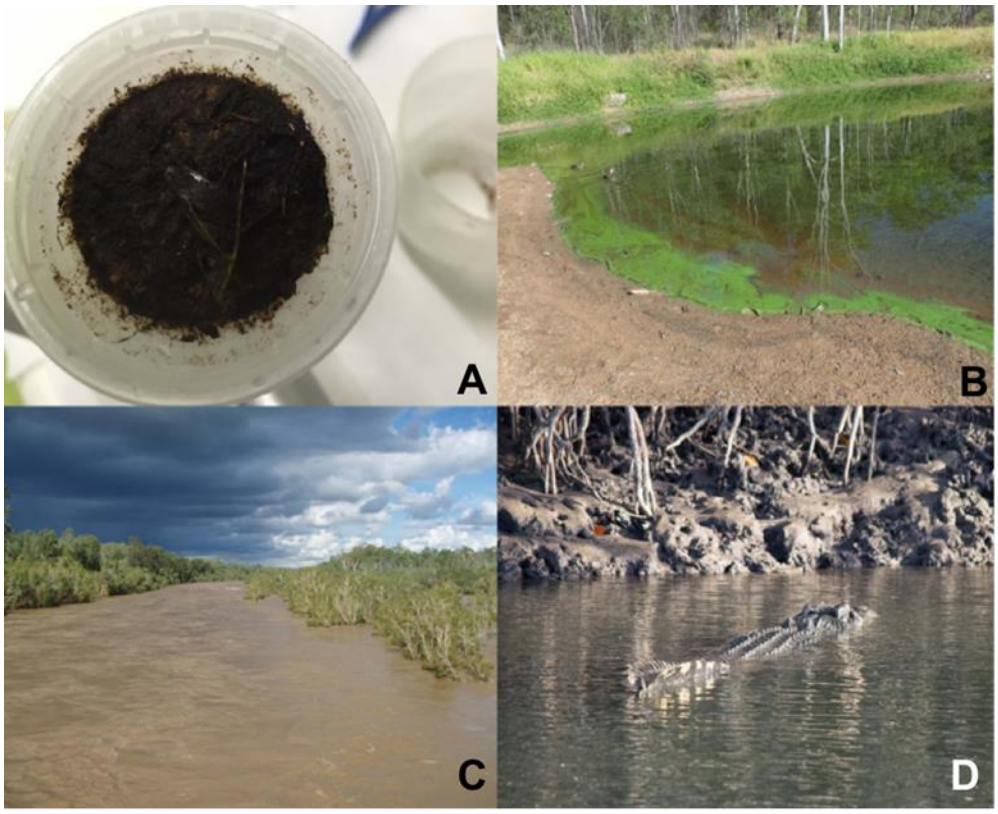
It has been suggested that elevated temperatures may accelerate the rate of eDNA degradation (Strickler et al. 2015). Indeed, Moyer *et al.*, (2014) showed that for every 1.02 °C increase in temperature, the per litre sample probability of eDNA detection decreased by 1.67 times. However, water temperatures of up to 35 °C were found to have no detectable effect on invasive Mozambique tilapia (*Oreochromis mossambicus*) eDNA degradation rates (Robson et al. 2016). Conversely, relatively high, but well within the range of tropical river systems, temperatures (35 °C) have been found to significantly increase fish eDNA shedding rates (Robson *et al*. 2016), which is likely due to increased metabolism or thermal stress, thus potentially positively affecting the detection probability of eDNA. Similarly, another study has indicated that fish release more eDNA in warm water than in cold water, and that eDNA concentration better reflects fish abundance/biomass at high temperatures (Lacoursiere-Roussel et al. 2016).

Exposure to high levels of ultraviolet radiation, particularly ultraviolet B (UV-B) light, can photochemically damage DNA, and aquatic environments at higher elevations or closer to the equator are more likely to experience increased effects of UV-B radiation on eDNA degradation rate (Strickler et al. 2015). It is however, likely that it is the interaction of multiple factors (pH, solar radiation, and temperature) either directly or mediated through the biological community, that influences the process of eDNA degradation in aquatic systems (Strickler et al. 2015).

High turbidity levels resulting from increased sedimentation and algal growth occur seasonally in tropical aquatic systems. These conditions present several challenges when collecting eDNA samples. Rapid clogging of filters (Figure 1.3A, B & C), requires the use of multiple filters per sample or filters with larger pore size, resulting either in increased filtration times or decreased capture rates of eDNA molecules. The presence of resuspended sediment in water samples may also affect the temporal scale of the data, as well as lead to downstream PCR inhibition of samples. Although eDNA in water reflects the present state of an ecosystem, eDNA can persist on the order of years in soils and sediments (Pedersen et al. 2015), thus mixing of contemporary and historic eDNA deposits could lead to misinterpretations as to the actual presence of a rare or invasive species in habitats of pressing environmental concern. Moreover, higher flow rates following wet season

precipitation events may lead to false negative detection due to longer than usual downstream transport distance of eDNA.

Several strategies may be adapted to tackle some of the challenges associated with eDNA studies in the tropics. These include: avoiding sampling during the wet season or right after heavy precipitation events; avoiding sampling during the hottest period of the day or during summer temperature highs; decreasing the risk of eDNA degradation by targeting areas/habitats that are less exposed (e.g., shaded, still) and filtering and preserving samples as soon as is practically possible. When sampling during heavy precipitation events cannot be avoided, increasing sample replication and sample volumes collected, as well as targeting samples from slow flowing or stagnant areas is recommended. To avoid rapid clogging up of the filtering apparatus and filters, filters with a larger pore size may have to be used. Filtering trials will have to be performed in order to identify the most optimal compromise between filtering constraints (e.g., filtering time and number of filters/sample) and eDNA capture probability for the species and environment of interest. For example, it was found that Mozambique tilapia (*Oreochromis mossambicus*) eDNA detectability decreased from 100% to 57% when using 3 μm and 20 μm filters, respectively (Robson et al. 2016). However, the significantly larger pore size decreased the filtering time from 44 min to 1.5 min per sample. In areas where dangerous wildlife co-occurs (e.g. crocodilians, Figure 1.3D), an extendable pole (Figure 1.2B) or a remotely operated sampling device (e.g., drone) may be used to collect samples safely. Lastly, PCR inhibition due to high levels of humic substances (e.g. humic and fulvic acid) in water samples is problematic. Strategies including use of PCR inhibition removal columns and/or dilution of samples should be applied in tropical aquatic systems.



1.7.8 **Reference databases**

Regardless of whether a species-specific barcoding or a community based metabarcoding approach is chosen, the reference database is a crucial starting (and end) point. When designing a species-specific qPCR assay, the reference database provides sequence information for target and exclusion species to ensure the specificity of the assay to the target species. In contrast, in eDNA metabarcoding approaches, the reference database is also used for taxonomic identification. If a specific species is present in a sample, but its barcode sequence is missing from the database, it will not be possible to identify it down to species level (but rather to genus or family) in a metabarcoding analysis.

It is essential to have a broad, accurate and curated sequence database. Genbank can be used as a starting point, but it may be necessary to collect samples from populations in the target region to build a bespoke reference database, appropriate for the sampling region. The Barcode of Life database may be of limited use, since it only contains COI sequences, which may not always be suitable for assay development. Certain gene regions may be too variable,

making it very difficult to design universal primers, or conversely, not variable enough, making it difficult to create species-specific qPCR assays or distinguish between different taxa in metabarcoding.

1.8 Future advances in eDNA (meta)barcoding

1.8.1 Factors influencing quantitative estimates

Both field and mesocosm/tank studies have shown that an increase in abundance or density of target species can lead to an increase in either eDNA concentration (Buxton et al. 2017; Klymus et al. 2015; Takahara et al. 2012; Thomsen, Kielgast, Lars L. Iversen, et al. 2012; Pilliod et al. 2013) or detectability (Eichmiller et al. 2014; Lacoursiere-Roussel et al. 2016; Mahon et al. 2013). In freshwater systems, using PCR and qPCR platforms, the rate of eDNA production has been positively correlated with biomass for several species, including common carp (*Cyprinus carpio*) in artificial ponds (Takahara et al. 2012), common spadefoot toads (*Pelobatus fuscus*) and great crested newts (*Triturus cristatus*) in natural ponds (Thomsen, Kielgast, Lars L. Iversen, et al. 2012), and tailed frog (*Ascaphus montanus*) tadpoles and both giant salamander (*Dicamptodon aterrimus*) larvae and paedomorphic adults (Goldberg et al. 2011; Pilliod et al. 2013). Because this relationship is not a clear-cut one, these studies suggest that eDNA can be used for relative rather than absolute quantification.

The amount of eDNA in the environment depends on both DNA shedding and degradation rates (which are dependent on a range of biotic and abiotic factors) (Dejean et al. 2011; Strickler et al. 2015), and heterogeneous dispersal of eDNA molecules via ecological processes (habitat specificity of target organisms) and/or presence of currents (Deiner & Altermatt 2014) or eddies (particularly in the marine environment). These factors are likely to vary seasonally in response to environmental changes and the life-history stage of the species in question (Barnes et al. 2014a; Buxton et al. 2017). For example, great crested newt (*Triturus cristatus*) eDNA concentration has been shown to increase within the breeding season due to reproductive behaviour and egg deposition, and subsequently with an increase of larval abundance (Buxton et al. 2017; Lacoursiere-Roussel et al. 2016). But it may also be associated with an increased eDNA production per biomass in juveniles, compared to adults, resulting from an increased metabolism during growth (Klymus et al. 2015). Likewise, it has been shown that seasonal variations in stream dwelling fish eDNA concentration were related

to total biomass (associated with breeding season and larval density), rather than abundance or behaviour (Doi et al. 2017). These factors may lead to either an over-or underestimate of organism density and as such, seasonal changes in eDNA concentrations may have implications for survey strategies, taking into account temporal and spatial patterns to target specific sampling windows, depending on the aim and species of the survey in question. Relating specifically to sharks, over the course of a year, there may for example be fluctuations in species densities and composition due to seasonal migrations related to water temperature (Guttridge et al. 2017; Kajiura & Tellman 2016; Kessel et al. 2016).

1.8.2 **Quantitative estimates using digital droplet PCR (ddPCR)**

Digital droplet PCR (ddPCR), also known as a "third-generation" PCR, provides a new method of sample analysis allowing for an accurate estimation of low concentrations of DNA. It has been suggested that ddPCR may be better suited for the detection of rare molecules in environmental samples compared to qPCR, and as such, it provides more accurate estimates of the abundance/biomass of a target species (Doi, Uchii, et al. 2015). Like standard PCR, ddPCR is a direct method that does not use calibration curves (derived from target DNA 'standards') to estimate target DNA concentration (Vogelstein & Kinzler 1999), thus decreasing the potential for user error (e.g. pipetting error when preparing standards or the introduction of contamination). Instead of a single measurement, the target DNA in ddPCR is randomly allocated into approximately 20,000 discrete droplets via microfluidics - some of which ideally contain only one or a few copies of the target DNA. The PCR occurs within each droplet, which is subsequently individually screened via fluorescence measurement for the presence of target DNA (Hindson et al. 2011; Pinheiro et al. 2012). Increasing the number of partitions (i.e. droplets) improves precision and therefore enables resolution of small concentration differences between nucleic acid sequences in a sample. With ddPCR it is possible to detect concentration differences between samples as low as 1.25-fold, which is more accurate than qPCR, which only allows for a 2-fold detection difference (Stoeckle et al. 2017; Hindson et al. 2011).

The ddPCR technique has been used to obtain absolute quantifications from a range of targets including a virus (Hindson et al. 2011), bacteria (Cave et al. 2016; Kim et al. 2014), fungi (Palumbo et al. 2016), and animal cells (Miotke et al. 2014), and more recently, a handful of fish species (Doi, Uchii, et al. 2015; Nathan et al. 2014; Jerde et al. 2016). Nathan

et al., (2014) were the first to use ddPCR to quantify the eDNA concentration of a fish species, the invasive round goby (*Neogobius melanostomus*) in mesocosm experiments. Comparing PCR, qPCR and ddPCR platforms, it was found that although both qPCR and ddPCR gave consistent estimates of DNA concentration, smaller variations in estimates were reported for ddPCR. Similarly, ddPCR proved to be more accurate in quantifying common carp (*Cyprinus carpio L.*) eDNA at low concentrations, than qPCR, suggesting this platform is very promising for use in estimating species biomass and/or abundance related to their eDNA concentration in the aquatic environment (Doi, Uchii, et al. 2015), once the species-specific relationship between biomass and eDNA concentration in the environment has been determined. In lotic systems, detection of eDNA is complicated by continuous dilution of the eDNA signal with simultaneous displacement downstream and/or mixing through physical processes (waves, currents). In a series of experiments in a semi-natural stream setting, it was concluded that at very low eDNA concentrations, there is an advantage to using ddPCR, as was demonstrated by the detection of target eDNA by ddPCR where qPCR failed (Jerde et al. 2016). Additionally, ddPCR has been shown to perform better in the presence of PCR inhibitors in field samples (Doi, Takahara, et al. 2015) compared to qPCR

1.8.3 **Quantitative estimates using metabarcoding**

A remaining controversial issue associated with eDNA metabarcoding is whether it can provide quantitative estimates; i.e., are the numbers of reads obtained for each species proportional to the abundance or the biomass of the species present in the original sample? Quantification of eDNA relating to species abundance could provide clues to habitat use, thus identifying spatial conservation priorities such as home ranges and dispersal and migration corridors (Barnes & Turner 2016). Although amplicon sequencing produces read counts that may contain valuable information about target species abundances (Evans et al. 2016; Port et al. 2016), the interpretation of the results of amplicon studies, in the context of quantitative ecology, is not straightforward and remains difficult (Kelly 2016). This is in part because the precise relationship between amplicon abundance and taxon abundance remains unknown and likely varies among taxa (Evans et al. 2016), as it is argued that PCR products are not fully proportional to real abundances due to the existence of primer bias (Clarke et al. 2014; Elbrecht & Leese 2015) and instead, some advocate for the use of PCR-free methods (Zhou et al. 2013). Consequently, the number of sequences obtained per taxon may currently not be

interpreted as quantitative but rather as semi-quantitative (Pompanon *et al.*, 2012; Kelly *et al.*, 2014)

For stream fish in lotic systems, a predictive model incorporating eDNA concentration has been developed in order to identify detection probabilities and abundance, as well as both eDNA production and degradation rates (Wilcox et al. 2016). Such models that include eDNA production, transport, and decay may improve the ability to infer organism abundance from eDNA quantity (Barnes & Turner 2016). As the relationships between eDNA and species abundance become clearer, the role of eDNA in estimating species abundance in both freshwater and marine environments is likely to become more valuable, increasing the potential of future eDNA applications in research and conservation.

1.8.4 **Increasing reference database coverage and taxonomic resolution**

Currently, the taxonomic resolution of sequences from eDNA metabarcoding datasets often does not reach the species level. Moreover, taxonomic misidentification poses a significant problem. One of the main causes is the incompleteness of reference databases. Taxonomic resolution may be increased, while simultaneously decreasing misidentification, by creating and updating a locally curated reference database. Moreover, when using group-specific primers, taxonomic resolution may be improved by complementing the primers with one or several additional primer pairs specifically designed to amplify more discriminately genetic regions for families with many closely related species (Valentini et al. 2016), such as is the case for the elasmobranch family of Carcharhinidae (Bakker et al. 2017). Additionally, broad spectrum primers often amplify non-target groups/species. This may potentially be overcome by the use of blocking primers (Vestheim & Jarman 2008), a strategy where the amplification of undesired sequences is specifically blocked.

1.8.5 **Taking eDNA analysis into the field**

A current limitation on the range and duration of eDNA field work is posed by the need to keep water samples chilled to prevent DNA degradation and ship them back to a central laboratory for processing. However, there are several recent developments that allow eDNA assays to be taken from the lab into the field. This allows for rapid detection of species or

even communities in the field, extends the range and duration of field trips, and takes away the need to ship samples to a laboratory. To enable field eDNA analysis, there are three critical steps that need to be made field capable. First, instead of shipping water samples to a central lab for filtration, the development of a mobile pump system allows for the filtration of water as the sample is being collected in the field (Laramie et al. 2015) (Figure 15.2). Field filtration in itself already simplifies field collection, removes the need for shipping large volumes of water, and improves sample preservation. Filters can be preserved in ethanol (Laramie et al. 2015), silica beads (Bakker et al. 2017) or modified buffers such as Longmire's buffer (Renshaw et al. 2015). Next, magnetic bead (Tomlinson et al. 2005) or syringe based (e.g. eDNA Water Filter Sample Prep Kit, Biomeme, Philadelphia, USA; Sterivex, Millipore, Inc., MA, USA) extractions remove the need for an immobile centrifugation step that is restricted to the lab. The final step, the analysis of the samples, can either be, a qPCR for a species-specific barcoding assay, or high-throughput sequencing for a metabarcoding assay. Improvements in miniaturisation and the use of mobile phones as small but powerful computing units have allowed the development of mobile qPCR thermocyclers such as the Two3, which can run a qPCR analysis for three samples in parallel using up to two fluorophores (Figure 1.4A). However, the increased field capability comes at the cost of reduced throughput. For metabarcoding, mobile high-throughput sequencing platforms such as the Oxford Nanopore MinION, which are used in conjunction with a laptop allowing for immediate analysis of the data, can be used to assess whole communities in the field (Figure 1.4B). As these two technologies advance, the throughput of these systems is expected to increase further.

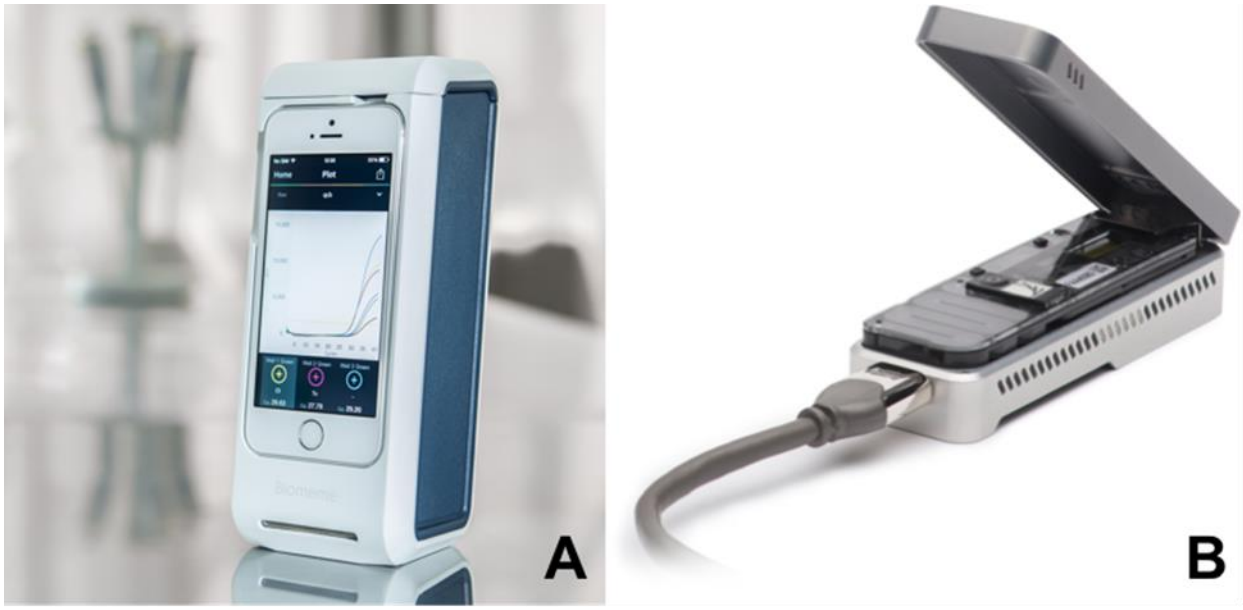


Figure 1.4 (A) Two3™ mobile qPCR thermocycler (Distributed by Biomeme) (**B**) MinION mobile high-throughput long-read sequencer (Oxford Nanopore)

1.8.6 **The emergence of autonomous sampling and analysis**

Applying remote and autonomous sampling techniques for eDNA collection may greatly expand the potential of eDNA applications to inform and improve conservation efforts (Barnes & Turner 2016). Hydroplane drones such as employed by Valentini *et al.*, (2016) (Figure 1.5A) are a promising way to collect large amounts of surface water with relatively little effort, as they can continuously filter water over an entire body of water and across areas otherwise difficult to sample. The increased sample size may enhance the detection probability of rare species (Hoffmann et al. 2016).

In the past couple of years, robotic systems, also referred to as 'ecogenomic sensors', for the autonomous collection and molecular analysis of eDNA samples have been under continuous development and are already being used by marine microbiologists to study marine microbial behaviour and to detect changes in bacterioplankton communities by utilising DNA probe (qPCR) and protein arrays to detect target molecules indicative of species and substances they produce (e.g. algal toxins) (Ottesen 2016; Preston et al. 2011; Scholin 2010). Additionally, autonomous high-resolution sampling, and both in situ and ex situ molecular essays have been used to study zooplankton distribution (Harvey et al. 2012). And a variety of molecular essays, including qPCR, have been applied in situ on particulates filtered from seawater from depths up to 4000 m (Ussler et al. 2013). These robotic

instruments (Figure 1.5B) are designed to autonomously collect, filter and analyse water samples from (sub)surface waters and, in near real-time, transmit data back to shore. In addition, they collect data on a wide variety of associated environmental parameters, such as currents, turbidity, salinity, and oxygen concentrations (Ottesen 2016). They may also preserve and archive water samples for laboratory analysis after the instrument is recovered (Breier et al. 2014; Scholin 2010).

Field portable molecular analytical techniques such as eDNA (meta)barcoding are still very challenging to implement in the context of remote instrumentation due to the requirement for multiple wet-chemistry processing steps (including concentration, extraction, and purification of the target DNA, followed by amplification) (Ottesen 2016). Hence, post-hoc macrobial eDNA analysis of microbe-motivated samples from these systems likely represents the first step towards broadening their use across taxa and disciplines (Barnes & Turner 2016). However, the use of ecogenomic sensors, specifically for the large-scale collection and analysis of macrobial eDNA, could bring about significant advances for molecular ecological studies and for the potential of eDNA applications to benefit conservation. As the range of deployable science instruments increases and their operating costs decrease, ecogenomic sensors will become an increasingly important tool for both oceanographic and ecological research. It will become possible to remotely monitor the presence, biodiversity and potentially abundance, of any marine species (including elasmobranchs) through cost-effective, long-term, high-frequency eDNA sampling regimes of the water column; including remote and inaccessible areas, such as the deep-sea.

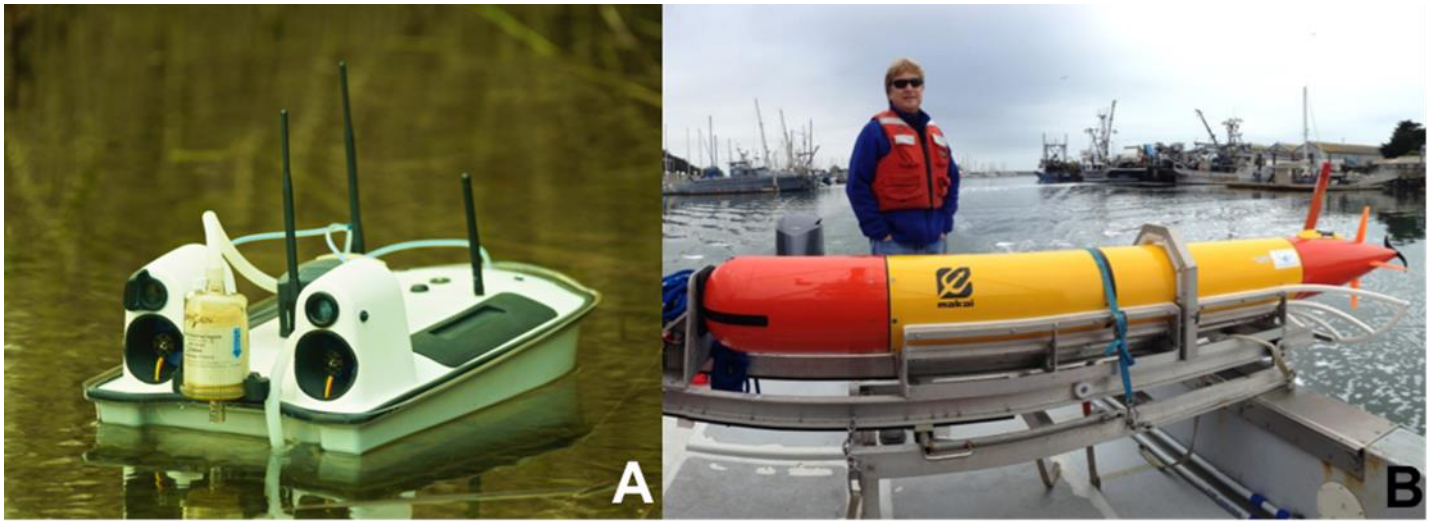


Figure 1.5 (A) Hydroplane drone-assisted water sampling for eDNA metabarcoding. This drone is double-hulled and the outer hull is disposable, which minimizes the risk of water body cross-contamination. Photograph credit: Alice Valentini and Tony Dejean/SPYGEN (**B**) The Environmental Sample Processor (ESP), a robotic microbiology laboratory that can filter water samples and either preserve the filtrate until recovery, or process the filtrate autonomously using a variety of molecular-probe techniques. Photograph credit: Todd Walsh © MBARI 2017. Taken from http://www.mbari.org/

1.8.7 **The use of long-range PCR for eDNA applications**

As DNA possesses limited chemical stability (Lindahl 1993), once it is shed from an organism, it will start to degrade. Hence, the general assumption regarding eDNA presence in an environmental sample, is that most eDNA will be highly degraded upon capture (Bohmann et al. 2014). Additionally, it has been shown that in most cases, recently released eDNA becomes undetectable within hours to days, underlying its usefulness in providing approximate real-time data on species presence in the environment. Coupled with current sequence length limitations of both qPCR and high-throughput sequencing platforms, most eDNA research has been focussed on a short fragment PCR amplicon sequencing approach to characterise macro-organismal species richness (Deiner et al. 2017; Olds et al. 2016; Valentini et al. 2016). Yet, one of the major drawbacks of targeting short eDNA fragments, is that it often limits the utility for species-level assignment (Deiner et al. 2016; Port et al. 2016). This can particularly be hampered when closely related species are concerned (Bakker et al. 2017), due to the highly conserved target sequences, between these species.

However, it has been shown that the largest percentage of Common carp (*Cyprinus carpio*) eDNA detected in water samples, was from particles that ranged in size from 1 to 10 μm (Turner, Barnes, et al. 2014). This is consistent with the presence of intact tissues or cells in aquatic environments, indicating that not all eDNA in a water sample is degraded. These findings are corroborated by earlier research on the detection of microbial genetic materials in the environment, which recognised that eDNA was present in both intracellular and extracellular forms (Ogram et al. 1987). It is likely that multicellular organisms shed genetic material into their environment first as sloughed tissues and whole cells, which subsequently break down and release DNA into the environment (Barnes & Turner 2016), and that consequently eDNA represents a complex mixture of particles ranging from extracellular DNA molecules up to whole cells and aggregations of cells (Turner, Barnes, et al. 2014). This suggests that eDNA for species currently occupying a habitat is not primarily free DNA suspended in solution, but that it could also be cellular or membrane bound DNA in a coiled or circular state, with comparatively more structural resistance to rapid degradation (Deiner et al. 2017; Torti et al. 2015; Turner, Barnes, et al. 2014).

As opposed to standard PCR amplification, long-range PCR makes it possible to produce a fragment that encompasses an entire mitochondrial genome (mitogenome) in a single amplification (Zhang et al. 2013). Hence, a recent study had set out to test whether it is possible to amplify and sequence entire fish mitogenomes from eDNA isolated from water

samples by applying long-range PCR amplification coupled with shotgun sequencing techniques (Deiner et al. 2017). By recovering full-length mitochondrial genes (COI, cytochrome B, 12S and 16S) this study has demonstrated that some of the eDNA from macro-organisms, currently inhabiting a water body remains intact for a short period, at least at the mitochondrial genome size.

However, one drawback of the method used in this study, is that mitogenome PCR products require shearing in order to fragment them prior to sequencing (due to the current read-length restrictions of the sequencing technology), and consequently is dependent on short fragment based de novo assembly or reference mapping (remapping the short reads to a reference sequence). This could still be an obstacle for the identification of closely related species, as conserved regions with high-sequence similarity are difficult to accurately assemble from a complex mixture (Deiner et al. 2017). It is therefore expected that with the continued advancement (e.g. improved cost-effectiveness and reduced error rates) of single molecule and long-read technologies, such as the Oxford Nanopore MinION (Laszlo et al. 2014) (Figure 1.4B), it will become possible to couple long-range PCR amplification and sequencing without fragmentation, avoiding problems associated with the use of short fragments.

Being able to sequence whole mitogenomes from eDNA, instead of having to rely on short fragment PCR amplifications for species identification, could potentially bring about major advances in taxonomic assignment; full-length barcodes, such as the COI region for animals (Hebert, Ratnasingham, et al. 2003), could be recovered in its entirety and used for species identification and additionally for the investigation of community structure and biodiversity. Future advances in long-read sequencing are expected to further advance eDNA applications into the realm of population and conservation genetics, systematics and phylogeography (Deiner et al. 2017).

Aims and objectives:

The detection and assessment of species and biodiversity trends is essential for the effective management and conservation of biodiversity, yet there is a lack of efficient and reliable tools to do so, particularly in the marine environment. Environmental DNA allows for the monitoring of marine species without requiring the collection of living organisms, and at much larger spatial scales, higher resolution and with more efficiency and replicability compared with more traditional sampling methods. Accordingly, this thesis focusses on the development and application of environmental DNA approaches in a range of natural marine environments. We set out to develop operational tools to assess marine biodiversity through eDNA analysis and to specifically address the following objectives:

- Assess the potential of eDNA for the detection of elasmobranchs, and subsequently further develop and apply the eDNA metabarcoding approach to assess and monitor elasmobranch biodiversity in natural marine environments, linking community differences detected by eDNA, with specific environmental and/or anthropogenic factors in both Atlantic and Pacific locations (**Chapter II**).
- Examine the performance of eDNA analysis in the detection and diversity assessment of elasmobranchs, in both impacted and wilderness areas, compared to traditional survey methods in New Caledonia (**Chapter III**).
- Evaluate the application of eDNA metabarcoding, using water samples from a range of Caribbean marine environments, for whole marine eukaryotic community scanning, and validating its potential to detect community differences related to geographic location and habitat type (**Chapter IV**).
- Conduct a comparative study, to investigate the performance of several metabarcoding primer sets for the diversity assessment of teleost fish communities, in coastal and transitional waters in the United Kingdom (**Chapter V**).

2 Chapter II

Environmental DNA reveals tropical shark diversity in contrasting levels of anthropogenic impact

Judith Bakker, Owen S. Wangensteen, Demian D. Chapman, Germain Boussarie, Dayne Buddo, Tristan L. Guttridge, Heidi Hertler, David Mouillot, Laurent Vigliola & Stefano Mariani

Author contributions: S.M. and J.B. conceived, designed and coordinated the study; J.B., S.M., D.D.C, G.B., D.B., T.L.G., H.H., D.M. and L.V. contributed to fieldwork and sample collection; Laboratory experiments and data analyses were conducted by J.B. and O.S.W.; J.B. wrote the manuscript; all authors read and commented on the manuscript.

2.1 Abstract

Sharks are charismatic predators that play a key role in most marine food webs. Their demonstrated vulnerability to exploitation has recently turned them into flagship species in ocean conservation. Yet, the assessment and monitoring of the distribution and abundance of such mobile species in marine environments remain challenging, often invasive and resource-intensive. Here we pilot a novel, rapid and non-invasive environmental DNA (eDNA) metabarcoding approach specifically targeted to infer shark presence, diversity and eDNA read abundance in tropical habitats. We identified at least 21 shark species, from both Caribbean and Pacific Coral Sea water samples, whose geographical patterns of diversity and read abundance coincide with geographical differences in levels of anthropogenic pressure and conservation effort. We demonstrate that eDNA metabarcoding can be effectively employed to study shark diversity. Further developments in this field have the potential to drastically enhance our ability to assess and monitor elusive oceanic predators, and lead to improved conservation strategies.

2.2 Introduction

Oceanic ecosystems are increasingly impacted worldwide. Marine predators are under often unsustainable fishing pressure, which has resulted in several documented cases of stock collapses (Jackson et al. 2001; Mullon et al. 2005; Myers & Worm 2003). Elasmobranch (sharks and batoids) populations specifically have suffered from overexploitation and stock declines (Camhi et al. 2009; Robbins et al. 2006; Simpfendorfer et al. 2002; Spaet & Berumen 2015; Ward-Paige et al. 2010; Worm et al. 2013). They are key species in virtually all marine trophic webs (Heupel et al. 2014; Navia Andrés F., Mejía-Falla Paola A., López-García Juliana, Giraldo Alan 2017) and have long been in conflict with human societies, due to their perceived competition with fishers (Gilman et al. 2008) or hazardous nature (Muter et al. 2013; Simpfendorfer et al. 2011). Owing to their relatively slow growth rate and low fecundity (Stevens 2000), they are also particularly vulnerable to overfishing (Bonfil 1994; Garcia et al. 2008; Musick et al. 2000). Only recently have elasmobranchs become the focus of conservation initiatives (Camhi et al. 2009; Dulvy et al. 2014), as the importance of these charismatic animals for the maintenance and resilience of healthy ecosystems is widely acknowledged (Baum & Worm 2009; Estes et al. 2011; Ferretti et al. 2010; Heithaus et al. 2012).

The development of management strategies for elasmobranchs depends on accurate population assessments in the field. Yet, currently established survey methods, such as fishing by long-lining or gill-netting, acoustic monitoring, baited remote underwater video (BRUV), underwater visual census (UVC) and fisheries-dependent population surveys, are often resource intensive, selective and dependent on taxonomic expertise, and sometimes invasive and potentially traumatogenic (Lodge et al. 2012; Simpfendorfer et al. 2016; Wheeler 2004). Therefore, biologists and managers worldwide are faced with considerable challenges due to the high effort and cost associated with the assessment and monitoring of elasmobranch biodiversity, abundance and distribution.

Environmental DNA (eDNA), DNA isolated directly from environmental samples such as soil or water, can be amplified, sequenced and assigned back to its species of origin through (meta)barcoding and has been suggested as an alternative to track species presence and abundance in their environment (Ji et al. 2013; Taberlet et al. 2012). Due to its limited persistence in the water column, in seawater even small (100-bp) eDNA fragments degrade beyond detectability within days (Thomsen, Kielgast, Lars Lønsmann Iversen, et al. 2012), the detection of eDNA from a specific taxon indicates its presence or very recent presence in

the environment (Barnes et al. 2014b; Jerde et al. 2011; Pilliod et al. 2014). Accordingly, over the past couple of years eDNA methods have increasingly been applied for the detection of rare and invasive species (Takahara et al. 2013; Turner et al. 2015; Wilcox et al. 2013). Moreover, it has been demonstrated that eDNA metabarcoding has the ability to outperform traditional survey methods for diverse taxa, including teleost fish, both in freshwater (Civade et al. 2016; Deiner et al. 2016; Hänfling et al. 2016; Valentini et al. 2016) and in marine ecosystems (Port et al. 2016; Yamamoto et al. 2017; Sigsgaard et al. 2016).

The first reported study to detect elasmobranch eDNA in natural water samples employed DNA barcoding (aiming to detect a single species in the environment), for the detection of the largetooth sawfish (*Pristis pristis*) (Simpfendorfer et al. 2016). Similarly, a species-specific approach was recently applied to amplify whale shark (*Rhincodon typus*) eDNA from oceanic water samples (Sigsgaard et al. 2016). On the other hand, eDNA metabarcoding has the potential to simultaneously identify several taxa from an environmental sample (Taberlet et al. 2012), which is clearly essential for community-level assessments.

However, previous studies have encountered challenges concerning elasmobranch specific detection, when applying this multispecific approach (Kelly, J. A. Port, et al. 2014; Miya et al. 2015). And although three species of elasmobranch have recently been detected in a large-scale marine eDNA study using a primer set designed for teleosts (Thomsen et al. 2016), we are still lacking evidence that eDNA metabarcoding can successfully be applied to describe elasmobranch diversity across a range of natural settings, for the purpose of ecosystem assessment and management.

Here, for the first time, we employ eDNA metabarcoding of natural seawater samples to specifically investigate shark communities in Atlantic and Pacific tropical ecosystems, using a previously published primer set targeting a 127 bp stretch of the mitochondrial COI region (Fields et al. 2015). We assess the potential of this low-effort approach for multi-species elasmobranch detection, and specifically examine whether patterns of species diversity and eDNA read abundance, reflect the known degree of anthropogenic impact in two independent tropical marine systems.

In the greater-Caribbean, there has been a long and ongoing history of elasmobranch exploitation, and high anthropogenic pressure in coastal zones has led to the broad-scale depauperation of elasmobranchs on Caribbean reefs (Ward-Paige et al. 2010; Newton et al. 2007). However, many species do still occur in populated areas where strong fishing regulations are in place or where specific shark conservation policies have been enacted

(Ward-Paige et al. 2010). In The Bahamas for instance, gillnet and long-line fishing have been prohibited since 1991 and their national waters have been declared a shark sanctuary in June 2011, prohibiting directed fishing or even the retention of shark by-catch (Chapman et al. 2013). In the wider Indo-Pacific region, overfishing and poaching are also responsible for declines in elasmobranch populations (Werry et al. 2014). Nevertheless, elasmobranchs do still occur in relatively high numbers around remote, isolated locations such as coral reefs on uninhabited atolls in the northern Line Islands (Sandin et al. 2008) and the Chagos Archipelago (Graham & Mcclanahan 2013). Furthermore, although several widely-distributed elasmobranch species are in effect cosmopolitan or circum-tropical, significant biogeographical differences exist between the Caribbean and the Pacific Coral Sea, which allows also for a broad-scale eDNA comparison of community composition.

Elasmobranch species inventories and assessment of geographical distributions based on eDNA metabarcoding could potentially represent an important tool for rapid environmental monitoring and hence influence conservation management and policy decisions. This study represents the first targeted effort that demonstrates the effectiveness of an eDNA metabarcoding approach for the detection and monitoring of elasmobranch communities.

2.3 Results

2.3.1 eDNA detection of elasmobranchs

A total number of 2,972,832 reads was obtained from an Illumina MiSeq run of pooled amplicon libraries, built from 55 Caribbean and 22 New Caledonian samples (Fig. 2.1 a, b and Supplementary material 2.1). A large part of the sequenced reads (80%) originated from non-specific amplification and were shorter than the target length. After sample assignment, quality and sequence-length filtering, 284,252 reads were left; of which 21,542 could be taxonomically assigned to elasmobranchs (Supplementary material 2.1). The number of elasmobranch reads per sample ranged from 0 to 5,205 (Supplementary Table S1). After the removal of singletons (MOTUs in a sample that contained only one read), taxonomic assignment from the sampled locations (Fig. 2.1a, b) resulted in 22 elasmobranch molecular operational taxonomic units (MOTUs), of which 12 were detected in the Caribbean, 16 in New Caledonia and 9 in both locations. Krona-like plots (Fig. 2.1c) display the complete taxonomic assignment for each of the sampling locations, while a Principal Component

Analysis (PCA) (Fig. 2.1d) depicts the scattering of all the samples containing elasmobranch reads across the two biogeographic areas. Even though several MOTUs are shared between the two regions, there is still a clear spread in MOTU (species) composition between New Caledonia and the Caribbean. No elasmobranch reads were detected in the two PCR negative controls that were performed and sequenced to detect potential contamination.

Using the 127-bp COI fragment, we did not find a wholly unequivocal correspondence between MOTUs and species; since some MOTUs had 100% sequence identity matches with more than one species in the BOLD database. Although this issue mostly pertained to the genus *Carcharhinus*, which is known to be taxonomically problematic and polyphyletic (Sorenson et al. 2014), it also affected the less speciose *Rhizoprionodon* and *Negaprion*. Consequently, this has also resulted in MOTUs containing sequences from both Coral Sea and Caribbean species that share an identical 127 bp sequence. Thus, for a more reflective separation, these particular MOTUs have been split into their Pacific and Caribbean components (see Supplementary material 2.2 for a complete data file of all reads per taxa, per sample).

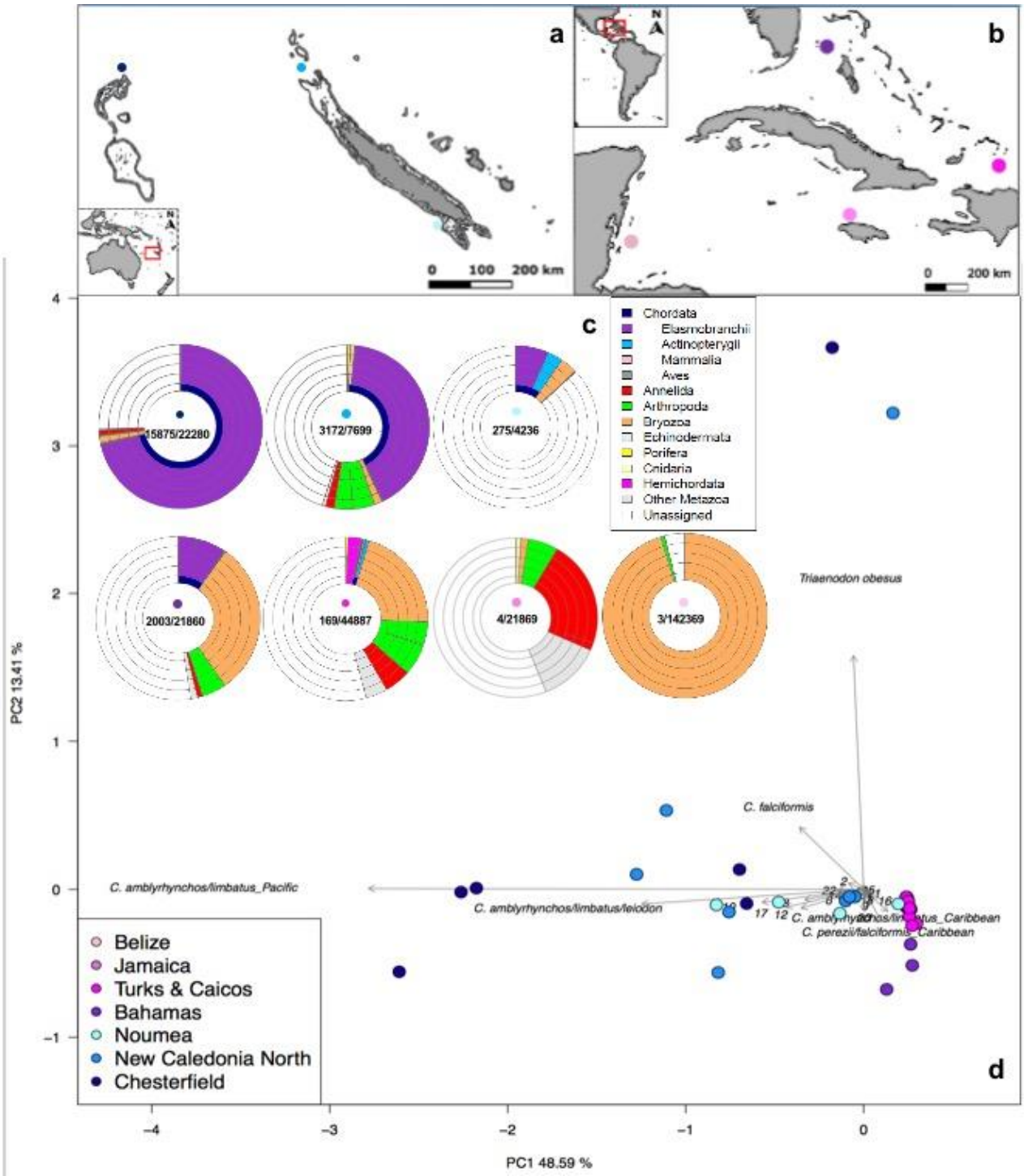


Figure 2.1 Map of New Caledonian (**a**) and Caribbean (**b**) sampling locations. The intensity of dot shading in all panels indicates the level of anthropogenic impact from 'severely impacted' (light pink/blue) to 'least impacted' (dark pink/blue). The krona-like plots (**c**) show the complete taxonomic assignment for each of the sampling locations (with elasmobranchs in purple). The different taxonomic levels are represented by the layers of rings, starting with phylum, for the innermost layer, and subsequently class, order, family and genus radiating outwards. In the centre of each location plot, the number of elasmobranch reads compared to the total number of filtered reads, is displayed. The Principal Component Analysis (PCA) (**d**) depicts the scattering of the samples containing elasmobranch reads, across the two biogeographic areas. The six most discriminating taxa are labelled in full, while the rest are indicated by numbers (following alphabetical order from lines 27-50 in Supplementary Table S2), namely: 1= *C. acronotus*, 2= *C. albimarginatus*, 4= *C. amblyrhynchos/limbatus_Caribbean*, 6= *C. brachyurus/perezii*, 8= *C. melanopterus/cautus*, 9= *C. leucas*, 10= *C. obscurus/macloti/longimanus/galapagensis*, 12= *C. perezii/falciformis_Pacific*, 13= *C. plumbeus*, 14= *C. plumbeus/altimus/sorrah*, 15= *G. cuvier*, 17= *N. brevirostris/acutidens_Pacific*, 18= *R. porosus/terraenovae*, 20= *S. mokarran*, 21= *D. Americana*, 22= *S. fasciatum*. Maps made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.

2.3.2 **Elasmobranch diversity and read abundance patterns**

The Bahamas is the Caribbean sampling location least subjected to fishing pressure, as a result of its shark sanctuary status; hence, it displays the greatest elasmobranch diversity, composed of 11 different MOTUs (Fig. 2.2). In the samples from the locations most impacted by anthropogenic disturbances, Jamaica and Belize, only 2 and 1 elasmobranch MOTUs respectively, were detected. A similar pattern is apparent for the New Caledonian samples, where the highest diversity is found in the most remote and pristine locations, the Chesterfield Atolls (11 MOTUs) and New Caledonia North (14 MOTUs). Contrastingly, only 5 elasmobranch MOTUs were detected in the capital, Noumea, the most densely populated area of New Caledonia.

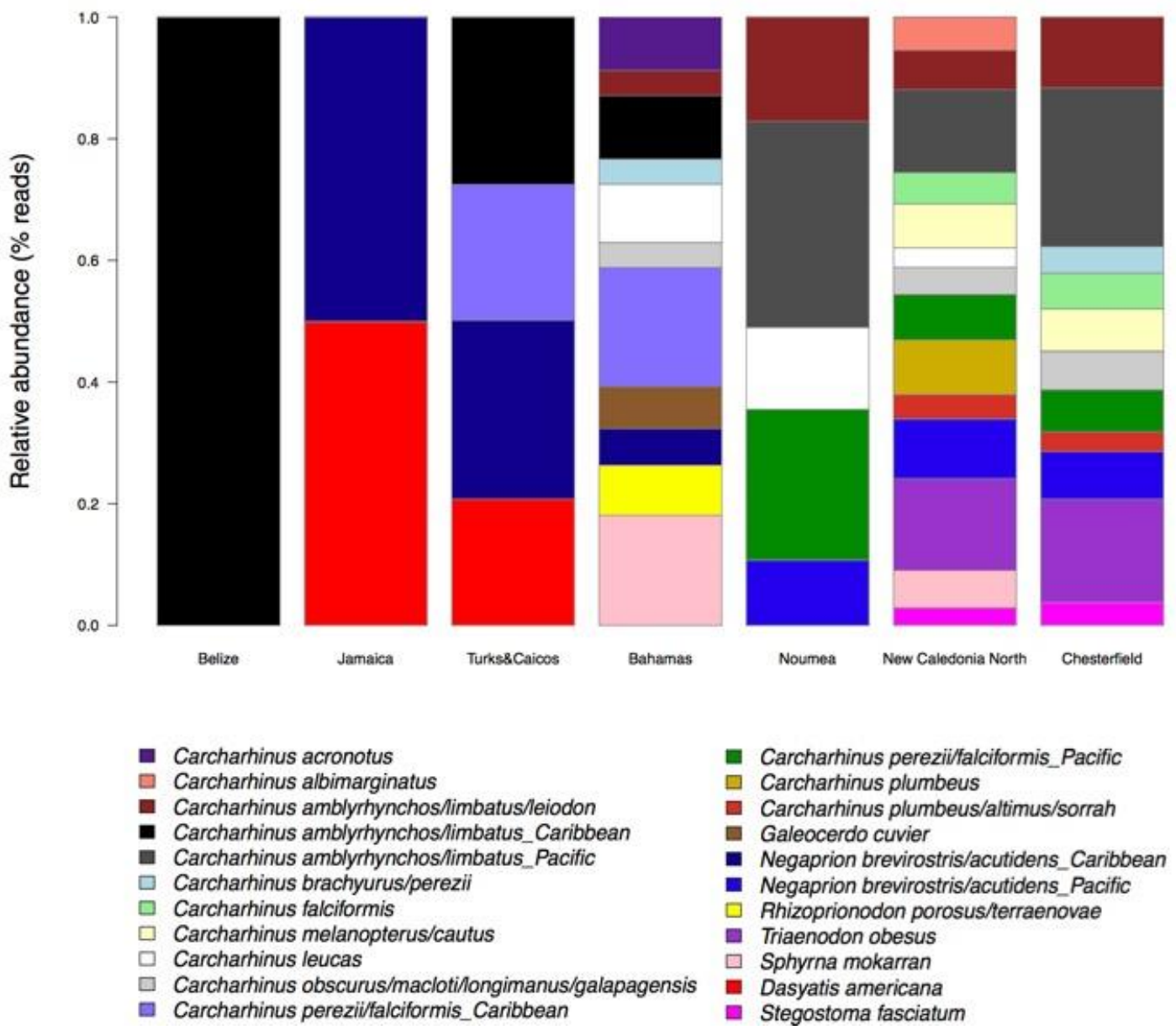


Figure 2.2 Bar plot showing the relative abundances of reads (fourth-root transformed) for every elasmobranch MOTU detected in the Caribbean and New Caledonian locations.

The violin plots of MOTU richness (Fig. 2.3a) and read abundances (Fig. 2.3b) show how the different sample values are distributed, by comparing the variable sample size distribution across the different locations. The distribution of density (number of MOTUs, Fig. 2.3a), and abundance of reads, (Fig. 2.3b) is represented by the width of the plots. For both the Caribbean (green) and the New Caledonian (blue) locations, MOTU richness (Fig. 2.3a) increases from left to right, following the pattern of decreasing anthropogenic disturbance. While the three least impacted locations, The Bahamas, New Caledonia North and the

Chesterfield atolls, show the greatest numbers of MOTUs in one sample (6, 7 and 8 MOTUs respectively), the violin plots for these locations show that richness is more equally spread across the different samples from Chesterfield (thus more samples containing multiple MOTUs); a sample is more likely to contain only 1 or 2 MOTUs in The Bahamas whereas in Chesterfield a sample is more likely to contain 4 or 5 MOTUs. Additionally, every sample from Chesterfield contains at least 3 elasmobranch MOTUs (for detailed MOTU richness per sample, see Supplementary Table S2). The abundance of reads (Fig. 2.3b) per location follows the same pattern: it increases in both Caribbean and New Caledonian locations, with a decreasing level of human impact. While both New Caledonia North and Chesterfield have samples that contain more than 1000 elasmobranch sequence reads, the number of samples with more than 1000 reads is greater in the more remote Chesterfield atolls. Additionally, Chesterfield is the only location without any samples with less than 80 sequence reads.

Generalized Linear Model (GLM) tests (using location as categorical factor) indicate significant differences in both diversity (DV) and read abundance (AB) between locations in both the Caribbean and New Caledonia. Pair-wise comparisons (post-hoc Tukey comparisons) show that for diversity (MOTU richness, Fig. 2.3a), significant differences are detected between The Bahamas and Jamaica ($P = 0.039$) and nearly significant differences between the Bahamas and Belize ($P = 0.095$). While diversity in New Caledonia is significantly different between Chesterfield and Noumea ($P = 0.013$). For read abundances (Fig. 2.3b), significant differences exist between The Bahamas and the other 3 Caribbean locations ($P < 0.001$). Additionally, read abundances in Turks & Caicos are significantly different from Jamaica ($P = 0.033$) and Belize ($P = 0.047$). In New Caledonia, all abundance comparisons are highly significant ($P < 10\text{-}10$).

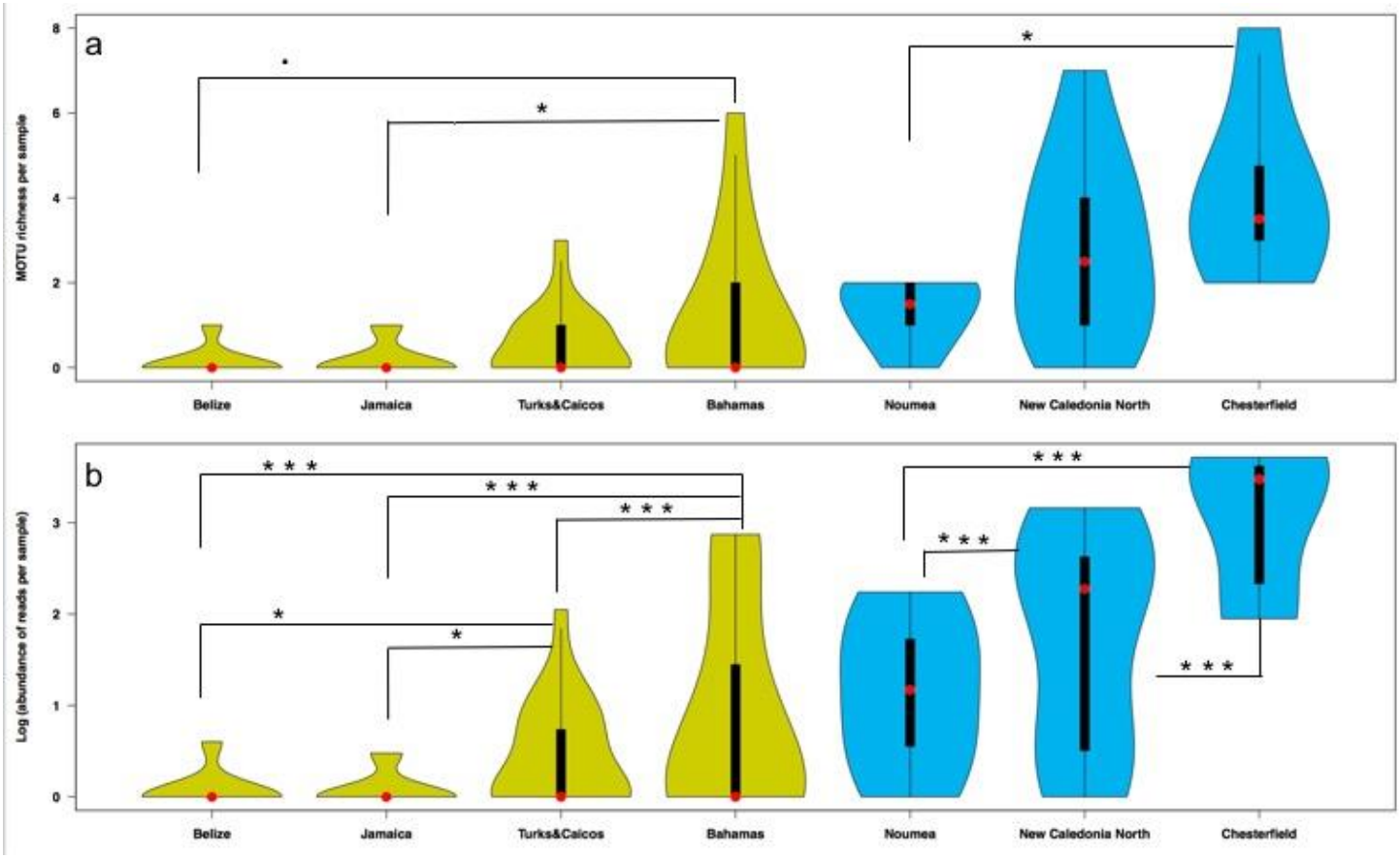


Figure 2.3 Violin plots showing (**a**) elasmobranch diversity (MOTU richness) and (**b**) abundance of reads per sample in the different locations from the Caribbean (green) and New Caledonia (blue). The shapes indicate the density distribution of the samples, extending from the minimum to the maximum observed values. The median values are indicated by the red dots. The thick black bars are the interquartile ranges. The thin black extending lines represent the 95% confidence intervals such that the values in the wider parts of the plots are more probable than those in the narrower parts. Per region, significant differences (P<0.05) are indicated with asterisks. Asterisk significant codes: *** P<0.001, **P<0.01, *P<0.05, • P<0.1

Species accumulation curves are plotted for each location (Fig. 2.4). The curves show elasmobranch diversity (MOTU richness) as a function of the number of samples in the locations from the Caribbean (A) and New Caledonia (B). Error bars indicate standard errors after 100 permutations. The results show that none of the Caribbean (Fig. 2.4a) or New Caledonian (Fig. 2.4b) samples tend to reach a plateau in MOTU richness, although, with the exception of The Bahamas, the Caribbean slopes tend to flatten after N=10. Non-saturation of species accumulation curves suggests that increased sampling effort would be desirable for capturing total diversity in each location, particularly in the naturally more diverse tropical Pacific.

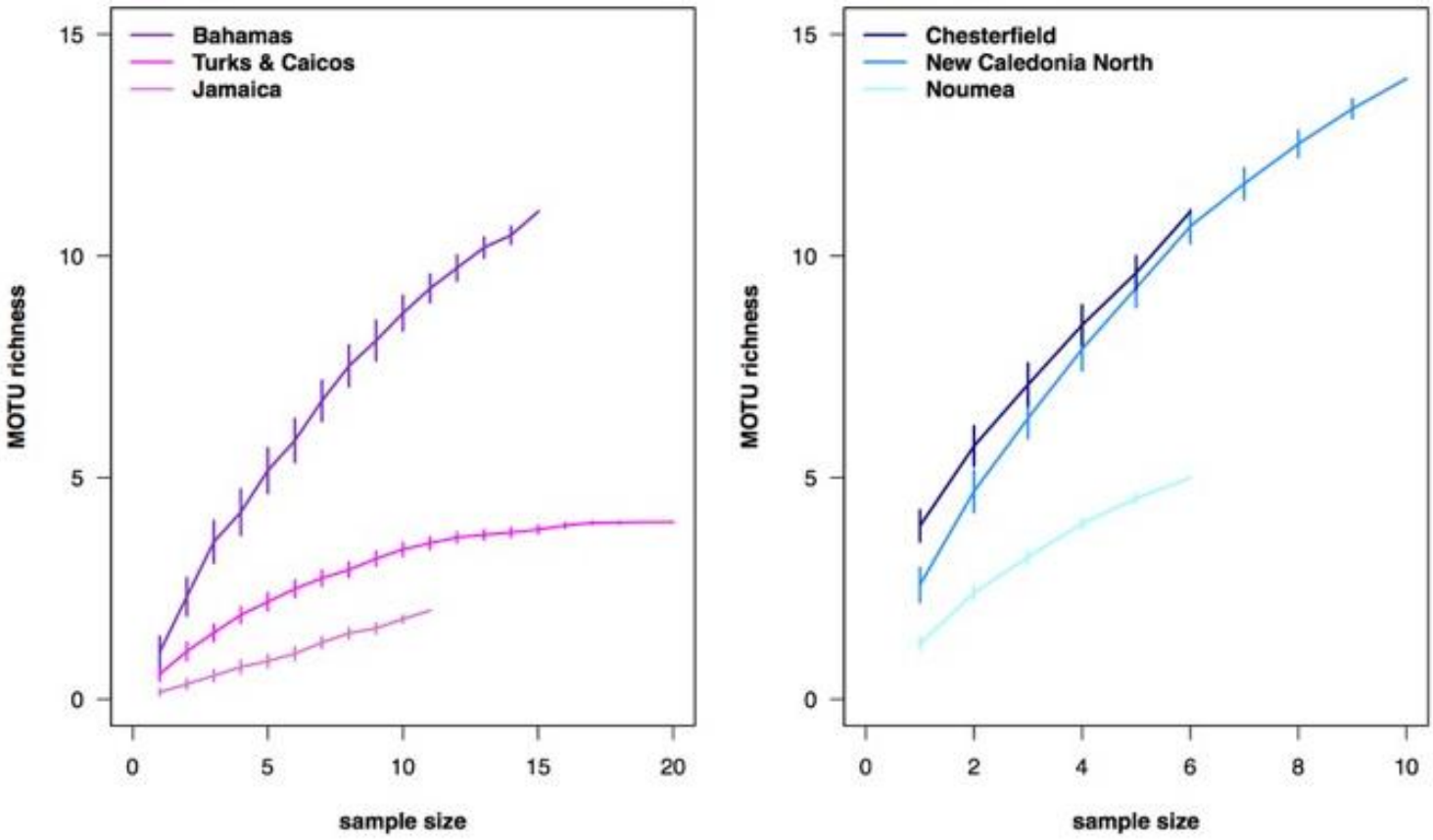


Figure 2.4 Species accumulation curves showing elasmobranch diversity (MOTU richness) as a function of the number of samples in the locations from the Caribbean (**a**) and New Caledonia (**b**). Error bars indicate standard errors after 100 permutations. Belize is absent from the plot as it contains only one elasmobranch MOTU.

2.4 Discussion

Our results demonstrate that eDNA metabarcoding can be applied to assess elasmobranch species richness and, potentially, relative abundance in natural seawater samples, the two main components of ecological communities. The derived geographical patterns of diversity and abundance of shark eDNA sequence reads may be used for monitoring purposes and ultimately to inform conservation management and policy decisions. At the global/macro-scale, the detected elasmobranch MOTUs collectively separate the Caribbean and New Caledonian regions (Fig. 2.1D), dominated by *Carcharhinus perezii* / *Negaprion brevirostris* and *Carcharhinus amblyrhynchos* / *Triaenodon obesus*, respectively, with the exception of the most depauperate locations (e.g. Belize, Jamaica and Noumea) which are grouped in the centre of the ordination plot. Additionally, the patterns of MOTU richness and abundance of sequence reads follow the level of anthropogenic impact in each location. Remote localities such as the Chesterfield atolls and protected areas such as Bimini, Bahamas show both the highest species richness and read abundance among our samples, whereas the less remote and non-protected locations show lower values for both diversity and abundance (Figs. 2.1C, 2.2 & 2.3).

In marine ecosystems, the impact on living resources is often framed into the Malthusian theory of human density around such ecosystems (Maire et al. 2016). Several studies have shown proximity to market to be the strongest predictor of overfishing on coral reefs (Cinner et al. 2012; Cinner et al. 2013; Cinner et al. 2006). A particular case study from New Caledonia has demonstrated that travel time from the market is a strong predictor of fish biomass, predator abundance and functional diversity on coral reefs (Maire et al. 2016; D'agata et al. 2016). Thus, remote locations such as the Chesterfield atolls, receive de facto protection due to their isolation (D'agata et al. 2016). The level of elasmobranch diversity and eDNA read abundance, sheds a new light on these wilderness areas which are already known to support high levels of fish biomass (Graham & Mcclanahan 2013; D'agata et al. 2016).

Likewise, it has previously been shown that sharks on reefs in the Greater Caribbean mostly occur in areas with low human population density or in a few places where strong fishing regulations or conservation measures have been implemented (Ward-Paige et al. 2010). While none of our Caribbean sampling locations is at more than 1 hour travel away from people, The Bahamas represents one of those locations whose elasmobranch populations in particular receive protection through effective conservation measures (Chapman et al. 2013).

Environmental DNA metabarcoding has a number of advantages compared to classical approaches for monitoring elasmobranchs. It is a minimally invasive and resource-effective technique. The eDNA sampling and metabarcoding protocols are easy to standardize and the molecular assignment does not require taxonomic expertise. Nevertheless, our findings also reveal a number of concerns that should be addressed in future developments of shark/elasmobranch eDNA metabarcoding approaches. First the taxonomic resolution of the final dataset is strongly dependent on the choice of markers; while the use of COI as a metabarcoding marker has previously been criticized (Deagle et al. 2014), owing to its high sequence variability, which may impair the design of truly universal primers and complicate bioinformatics analysis - it can also be argued that COI presents two major advantages over other potential markers. First, the steadily growing international effort, headed by the Consortium for the Barcode of Life (CBOL), to develop a public DNA barcoding database with curated taxonomy, greatly facilitates taxonomic assignment. The BOLD database (http://www.boldsystems.org/) (Hebert, Cywinska, et al. 2003; Ratnasingham & Hebert 2007) currently includes >4 million sequences belonging to over 500,000 species, curated and identified by expert taxonomists. Secondly, the high mutation rate of COI enables identification at the species level, whereas the highly conserved sequences of other markers, such as 18S, make it often impossible to distinguish at the species or genus levels; and species-level identification is crucial for studies aimed at detecting rare species, such as is often the case for sharks.

Nevertheless, using the 127 bp elasmobranch specific COI fragment (Fields et al. 2015), we still recorded some ambiguity in the taxonomic assignment of some species of the genera *Carcharhinus, Rhizoprionodon* and *Negaprion*, owing to the limited sequence variability within the amplicon. Consequently, the sequences of some MOTUs are 100% identical to individuals belonging to different species in the BOLD database. In our dataset, this is presented by several MOTUs belonging to either of two or more species. Furthermore, this has resulted in 3 MOTUs having 100% sequence identity for species occurring in New Caledonia, and species occurring in the Caribbean. One particular example is the Carcharhinus *amblyrhynchos*/*limbatus* MOTU. The bulk of sequences within this MOTU are from New Caledonia, 16730 compared to 146 in the Caribbean (Supplementary material 2.2). In all probability, the sequence reads from New Caledonia belong to both *C. amblyrhynchos*, the grey reef shark (a species abundant in this area and also the species most often visually detected during sampling operations, Supplementary material 2.1) and to *Carcharhinus limbatus*, the blacktip shark. Since *C. amblyrhynchos* does not occur in the Caribbean, the

146 reads of this MOTU in the Caribbean samples most likely belong to *C. limbatus*. Similarly, the *Negaprion brevirostris/acutidens* and *Carcharhinus perezii/falciformis* MOTUs are shared between New Caledonia (365 and 206 reads respectively) and the Caribbean (91 and 1045 reads respectively). *Negaprion brevirostris* (lemon shark) is native to the Americas whereas *N. acutidens* (sicklefin lemon shark) is widely distributed throughout the Indo-Pacific. *Carcharhinus falciformis*, the silky shark, is circumtropical, while the distribution of *Carcharhinus perezii* (Caribbean reef shark) is restricted to the tropical western Atlantic Ocean.

In order to resolve the issue with closely related species, it will be essential to design alternative primers that are able to amplify a longer fragment of the gene region, in addition to an improved reference database. In relation to this, it is clear that certain elasmobranch species present in the environment at the time of sampling could not be detected using the selected primer set, as their non-degenerate sequences contain mismatches with the binding regions of several species. This is epitomized by the case of the nurse shark, *Ginglymostoma cirratum*, which is an abundant species in the Caribbean and some individuals were visually observed at the time of sampling. Yet, eDNA sequence reads of *G. cirratum* were not detected in any of our Caribbean samples. Comparing the available *G. cirratum* COI sequences in public repositories with our primer sequences, it becomes apparent that two mismatches with the primer 3' end are most likely responsible for the prevention of amplification of this species' DNA from environmental samples. In silico mismatch statistics, comparing the 3' half of the primers with full mitochondrial genome sequences available, are listed in Supplementary material 2.4, for all elasmobranch Orders. This table shows that the primers are particularly suitable for amplifying Carcharhiniformes, which contains more than half of all shark species, including those of most ecological relevance for this study. However, all the elasmobranch orders may be amplified. Potential mismatch issues may be resolved by decreasing the specificity of the primer set by incorporating degenerate bases or inosine nucleotides. However, this approach may have the undesired effect of an increase in the number of reads belonging to non-target taxa getting amplified. Clearly, this trade-off between taxonomic resolution and non-target amplification needs to be well balanced prior to applying the eDNA metabarcoding approach for the purpose of informing conservation and management decisions. Here, we focused on large-scale differences and overall patterns in relation to anthropogenic influence, but in alternative contexts it may be necessary to attain greater taxonomic accuracy (e.g. endangered or invasive species); a challenge also faced by

currently established ribosomal amplicon-based analysis of other vertebrates (Thomsen et al. 2016; Kelly et al. 2016).

Quantification of eDNA relating to species abundance could provide clues to habitat use and preference, thus identifying spatial conservation priorities such as home ranges and dispersal and migration corridors (Barnes & Turner 2016). However, whether eDNA metabarcoding can provide quantitative estimates, particularly in the case of community-level abundance, remains a controversial issue. Although amplicon sequencing produces read counts that may contain valuable information about target species abundances (Port et al. 2016; Evans et al. 2016), the interpretation of the results of amplicon studies, in the context of quantitative ecology, is not straightforward and remains ambiguous (Kelly 2016). This is in part because shedding rates between communities, species and individuals may differ. But also because the precise relationship between amplicon abundance and taxon abundance remains unknown and likely varies among taxa (Evans et al. 2016; Elbrecht & Leese 2015), as it is argued that PCR products are not fully proportional to real abundances due to the fact that primer efficiency may vary among species templates (primer bias) (Elbrecht & Leese 2015; Clarke et al. 2014). While previous studies have shown positive rank correlations between species abundance and read abundance (Hänfling et al. 2016; Kelly, J. A. Port, et al. 2014; Evans et al. 2016; Klobucar et al. 2017), as of yet, no studies have been published, revealing evidence of a relationship between relative abundance of species within a community and their respective eDNA read abundances. And while currently, no experiments have been performed to empirically verify the relationship between read abundance and community biomass for elasmobranchs in particular, our data show that read abundances are higher in the more pristine/remote sampling locations and that these patterns of read abundance are coherent with expectations that can be inferred from the contrasting levels of human impact/remoteness of the different locations, for both the Caribbean and New Caledonia (Fig. 2.3B). While for example the species diversities of New Caledonia North and Chesterfield are very similar (Fig. 2.2), eDNA read abundance is significantly higher in Chesterfield (Fig. 2.3B), suggesting that read abundance may be correlated with remoteness. As the relationships between eDNA and species abundance become clearer, the role of eDNA in estimating species abundance in both freshwater and marine environments is likely to become more valuable, increasing the potential of future eDNA applications in research and conservation.

Conservation and management of elasmobranch diversity relies on the effective monitoring of species across large oceanic areas. While direct observation and identification

of individuals are often complicated, we have demonstrated that, despite the seemingly daunting task of probing vast stretches of ocean by collecting water samples, eDNA metabarcoding has great potential for developing into an objective and powerful elasmobranch assessment tool, applicable to a wide range of ecological goals, from the mapping of diversity gradients in response to environmental variation, to the monitoring of the effectiveness of spatial protection measures.

2.5 Material and methods

2.5.1 Experimental design

Aqueous eDNA samples were collected with interoceanic replication, to test for spatial marine protection as a predictor for elasmobranch diversity. During February and March of 2015 (Supplementary material 2.1), samples were collected from four Caribbean locations impacted by various levels of anthropogenic pressures (Fig. 1.1B). Jamaica is known to have one of the most depauperate fish populations in the Caribbean and a severely extirpated elasmobranch fauna (Hawkins & Roberts 2004); thus, it was expected that Jamaica would sit at the lower end of the elasmobranch diversity range and read abundance. In Belize sampling took place around the partially submerged Glover's Reef atoll, which is part of the Mesoamerican Barrier Reef. Even though this region has a relatively large number of marine reserves, including the 'Glovers Reef Marine Reserve', shark sightings in the Caribbean are quite rare and relatively few shark sightings occurred in the Mesoamerican Barrier Reef area during a previously conducted survey (Ward-Paige et al. 2010; Bond et al. 2012).

In the Turks & Caicos Islands, where sampling took place around South Caicos, the establishment of a shark sanctuary is under consideration; however, the islands are currently still experiencing high fishing pressure (Newton et al. 2007), which tends to disproportionally reduce densities of longer-lived, larger-bodied individuals (Sandin et al. 2008). At the other end, the nation of The Bahamas is a designated shark sanctuary (Chapman et al. 2013) and as such, an area characterised by consolidated shark protection. The sampling was conducted around the islands of Bimini, which consequently boast an abundant and diverse elasmobranch fauna (Guttridge et al. 2017; Jennings et al. 2012).

In the tropical Pacific, eDNA samples were collected from three locations in New Caledonia (Fig. 1.1a), during September, October and November of 2015 (Supplementary material 2.1).

New Caledonia has a unique anthropogenic impact gradient from nearly pristine to significant levels of anthropogenic disturbance (D'agata et al. 2016). The most heavily impacted site is represented by the capital Noumea, the most densely populated area in New Caledonia (Maire et al. 2016). However, most reefs near Noumea are no-take reserves and shark fishing is historically non-existent in New Caledonia so that near Noumea, shark populations may be healthier compared to many other impacted areas. Sampling sites north of the main island Grande Terre, 'New Caledonia North' represent the intermediate level of anthropogenic impact, being between 70-120 km removed from the nearest human settlement and a minimum of 500km/15 hours travel time, from the Noumea fish market. The isolated Chesterfield atolls, 550 km northwest of Grande Terre (~35 h travel time from the Noumea fish market), are the most remote of all our sampling locations. These samples were expected to show the highest levels of elasmobranch diversity and eDNA read abundance. Within all 7 (Caribbean and New Caledonian) locations, samples were collected from between 6 and 20 different sites covering a variety of habitats (Supplementary material 2.1 contains coordinates per site). A total of 55 samples from the Caribbean and 22 samples from New Caledonia were collected and analysed. Each sample consisted of 4 litres of sea water, collected by either a Kemmerer type water sampler or directly with a plastic collection bottle.

2.5.2 **Sample processing and DNA extraction**

After collection, the water samples were individually covered and stored, in the dark and on ice, during transport to the local laboratory facilities. Vacuum filtration was carried out within two hours after collection. When it was not feasible to carry out filtration within two hours after collection, due to travel time to laboratory facilities, the samples were directly frozen after collection, until further processing. The sterile mixed cellulose esters (MCE) filters (Merck Millipore; 47 mm diameter; 0.45 µm pore size) containing sample filtrates were stored in 2.0 ml screw-cap microcentrifuge tubes containing silica beads. The silica beads function as a desiccator, drying out the filters and hence preventing the DNA from degrading. The sample filters were then stored at -20°C until extraction. DNA was extracted from the filters with the Mo-Bio PowerSoil DNA Isolation Kit (www.mobio.com), following the manufacturers' protocol. Purified extracts were assessed for DNA concentration in a Qubit fluorometer (Thermo Fisher Scientific).

2.5.3 **Contamination control**

Contamination of samples may occur anywhere from preparing sampling equipment and collecting the samples in the field (target DNA being carried unintentionally from one locality to another), to every subsequent step of sample preparation, extraction and analysis in the laboratory. Hence, strict adherence to contamination control was followed at all field and laboratory stages in order to prevent the occurrence of contamination, including the use of disposable gloves and single use-sterile collection bottles and filtration equipment, and the bleaching (50% bleach) of sampling devices and laboratory equipment and surfaces. Additionally, a dedicated controlled eDNA lab at the University of Salford, with separate rooms designated for the physical separation of eDNA extraction, pre-PCR preparations and post-PCR procedures, was used for all laboratory work. Moreover, to identify potential contamination, DNA extraction blanks (elution buffer from extraction kit) and PCR blanks were included.

2.5.4 **Library preparation and sequencing**

For the amplification of eDNA metabarcoding markers, an elasmobranch specific COI primer set was used. This previously published primer set consisted of a novel reverse primer 'Shark COI-MINIR' 5'-AAGATTACAAAAGCGTGGGC-3' (Fields et al. 2015) and two universal fish barcoding forward primers FishF2 5'-TCGACTAATCATAAAGATATCGGCAC-3' and VF2 5'-TCAACCAACCACAAAGACATTGGCAC-3' (Ward et al. 2005), yielding an amplicon of 127 bp (Fields et al. 2015). For the multiplex Illumina sequencing run, we used 4 sets of 24 primers with attached 8-base sample-specific oligo-tags differing in at least 3 bases (Guardiola et al. 2015). In order to increase variability of the amplicon sequences, a variable number (2, 3 or 4) of fully degenerate positions (Ns) was added at the beginning of each primer (Wangensteen & Turon 2017). The full, sequenced PCR product, consisted then of 195 bp, including the amplicon, primers, sample tags and leading N's.

For PCR amplification, a single step protocol was used, directly attaching the 8-base tagged primers. The PCR mix recipe was as follows: a total volume of 20 µl included 2 µl 10x buffer (BioLine), 0.6 µl 50mM MgCl (BioLine), 0.5 µl of each of the 5 µM forward primers (Eurofins), 1 µl of the 5 µM reverse primer, 0.2 µl 10 mM dNTP mix (BioLine), 0.2 µl BioTaq DNA polymerase (5u/µl, BioLine), a standardised amount (10 ng) of the filter-

extracted eDNA template, and 13 µl sterile water. The PCR profile included an initial denaturing step of 95 ºC for 15 min, 35 cycles of 94 ºC 1 min, 52 ºC 1 min and 72 ºC 1 min and a final extension step of 72 ºC for 5 minutes. The quality of all amplifications was assessed by electrophoresis, running the products through a 1.5% agarose gel stained with Gel Red (Cambridge Bioscience) and visualized on a UV light platform. All PCR products (including one replicate per sample and 2 PCR negative controls) were pooled into 4 multiplexed sample pools (each composed of 24 individually-tagged samples) and purified using MinElute columns (Qiagen). Four Illumina libraries were subsequently built from the four pools, using the NextFlex PCR-free library preparation kit (BIOO Scientific). The libraries were quantified using the NEBNext qPCR quantification kit (New England Biolabs) and pooled in equimolar concentrations along with 1% PhiX (v3, Illumina) serving as a positive sequencing quality control. The libraries with a final molarity of 8 pM were sequenced on an Illumina MiSeq platform in a single MiSeq flow cell using v2 chemistry (2x150 bp paired-ends).

2.5.5 **Bioinformatic and statistical analysis**

The bioinformatic analysis was based on the OBITools metabarcoding software suite (Boyer et al. 2016). The pipeline used for data analysis is summarized in Supplementary material 2.3. Quality of the reads was assessed using FastQC. Paired-end reads were aligned using illuminapairedend and alignments with quality score >40 were kept. The aligned dataset was demultiplexed using ngsfilter. The length distribution of the demultiplexed reads showed a large percentage of short fragments (<95 bp), originating from non-specific amplifications and primer-dimer artefacts, which were not removed during the size selection step of library preparation. Thus, a length filter (obigrep) was applied to the aligned reads (120-135 bp) in order to select only the fragments with the correct target size. Reads containing ambiguous bases were also removed. The reads were subsequently dereplicated using obiuniq and a chimera removal step was performed using the uchime-denovo algorithm (Edgar et al. 2011) implemented in vsearch (Rognes et al. 2016). The MOTUs were delimited using the sumaclust algorithm (Boyer et al. 2016) with a constant similarity threshold of 99%. Taxonomic assignment of the representative sequences for each Molecular Operational Taxonomic Unit (MOTU) was performed using the ecotag algorithm (Boyer et al. 2016). We built a bespoke elasmobranch reference database using a custom R script for retrieving all

COI elasmobranch sequences available from the BOLD database (Ratnasingham & Hebert 2007), and subsequently selecting those that included our 127 bp target fragment. In order to add homologous sequences from other, non-elasmobranch taxa, an in silico PCR was performed against release R117 of the EMBL-EBI database using ecoPCR (Ficetola et al. 2010). Subsequently, the obtained reference sequences were added to the elasmobranch sequences obtained from BOLD. These additional reference sequences were added to our elasmobranch database in order to avoid the incorrect assignment of amplified sequences, belonging to other taxa, to elasmobranchs. This combined reference database is available from http://github.com/metabarpark/Reference-databases. The final refining of the dataset included taxonomy clustering of MOTUs assigned to the same species.

Due to its high sensitivity, an additional challenge associated with eDNA metabarcoding, is the risk of contamination (Goldberg et al. 2016; Thomsen & Willerslev 2015), and hence the possibility of introducing false positive results. While it is certainly possible to detect a species present in a sample, represented by a single sequence read, it is not possible to completely exclude contamination (or sequencing error) as the potential cause of MOTUs containing only a single read, i.e. to dismiss single reads as potential false positives. Accordingly, we have opted for a more conservative approach and have removed all single read MOTUs from our samples.

All statistical analyses were performed in R v 3.3.0 (https://www.R-project.org/). The vegan package v. 2.4-0 (Oksanen *et al*., 2016) was used for the calculation of sample-based species accumulation curves. A generalized linear model approach was used for testing differences in MOTU richness and read abundances (square-root transformed) as a function of location, using the glm2 package v. 1.1.2 (Marschner 2015). The Poisson distribution family function was used for modelling the residuals and package multcomp (Hothorn et al. 2016) was used for post-hoc Tukey comparisons. All custom R scripts are publicly available from http://github.com/metabarpark.

2.6 Acknowledgements

We thank all staff, students and volunteers from Glover's Reef Research Station, Discovery Bay Marine Laboratory and Field Station, Montego Bay Marine Park, Bimini Biological Field Station (Sharklab), the SFS Center for Marine Studies, and the crew of the research vessel Amborella, for their assistance with sample collection. We also thank Alexandre De Menezes and Ian Goodhead for technical discussions and Andjin Siegenthaler for assistance

with our sampling area maps. We are also grateful to Kathryn Matthews and Angelo Villagomez for logistic support. This work was co-funded by the Pew Charitable Trusts, the University of Salford R&E strategy funding, the Total Foundation and the Government of New Caledonia.

3 Chapter III

Environmental DNA illuminates the dark diversity of sharks

Germain Boussarie†, Judith Bakker†, Owen S. Wangensteen, Stefano Mariani, Lucas Bonnin, Jean-Baptiste Juhel, Jeremy J. Kiszka, Michel Kulbicki, Stephanie Manel, William D. Robbins, Laurent Vigliola‡, David Mouillot‡
† These authors contributed equally to this work as first authors. ‡ These authors contributed equally to this work as senior authors.

Author contributions: G.B., S.M., D.M. & L.V. designed the study with inputs from all co-authors. J.B., O.S.W. & S.M. designed and carried out the eDNA laboratory procedures and bioinformatics pipeline. M.K., L.V., J-B.J., G.B., W.R. and many others collected the UVC, BRUVS and eDNA samples. G.B., J.B., O.S.W., L.V. & D.M analysed the data and drafted the manuscript with input from all authors.

The work in this chapter is the result of a collaboration in which the data collected for this thesis has been combined with data collected by collaborating authors, in order to compare the results of eDNA methods with more traditional methods of elasmobranch sampling.

3.1 Abstract

In the era of 'Anthropocene defaunation', large species are often no longer detected in habitats where they formerly occurred. However, it is unclear if this apparent missing, or 'dark', diversity of megafauna results from local species extirpations, or from failure to detect elusive remaining individuals. We find that despite two orders of magnitude less sampling effort, environmental DNA (eDNA) detects 44% more shark species than traditional underwater visual censuses and baited videos across the New Caledonian archipelago (south-western Pacific). Furthermore, eDNA analysis reveals the presence of previously-unobserved shark species in human-impacted areas. Overall, our results highlight a greater prevalence of sharks than described by traditional survey methods, in both impacted and wilderness areas. This indicates an urgent need for large-scale eDNA assessments to improve monitoring of threatened and elusive megafauna. Finally, our findings emphasize the need for conservation efforts specifically geared towards the protection of elusive, residual populations.

3.2 Introduction

Human activities are largely responsible for the ongoing defaunation of ecosystems worldwide, causing massive population declines and local species extirpations (Dirzo et al. 2014; Young et al. 2016). This global wave of defaunation may dramatically increase local 'dark diversity', defined as the suite of species which should be present within a certain region, based on their habitat requirements and dispersal ability, yet, are absent (Pärtel et al. 2018). In other words, dark diversity encompasses the diversity of locally absent species, although biogeographic history, as well as ecological and environmental conditions, suggest their presence (Moeslund et al. 2017). High dark diversity may imperil ecosystem functioning (Soliveres et al. 2016), but also represents potential for recovery of purportedly absent species (Pärtel et al. 2018; Lewis et al. 2017). A substantial portion of megafaunal dark diversity is composed of mobile, rare, elusive, and threatened species that are highly challenging to detect (Dirzo et al. 2014). Accordingly, an ongoing concern is whether this megafaunal dark diversity has been correctly measured, or overestimated as a result of non-detection of remaining individuals by traditional sampling methods. The answer to this question has significant implications in terms of management and conservation, as the presence of previously undetected individuals may require immediate action to prevent extirpation of remnant individuals, while the confirmed absence of species requires different management considerations (Lewis et al. 2017).

Detecting species occurrences and extirpations is more challenging in the ocean than on land, since most habitats remain hardly accessible and therefore poorly investigated (Webb & Mindel 2015). Similarly, accurate assessment of dark diversity is particularly problematic for low density, mobile species such as sharks. Sharks are one of the most threatened marine taxa (Davidson & Dulvy 2017). They often have a high intrinsic vulnerability to fishing due to slow population growth (Hobday et al. 2011) and with shark products such as dried fins reaching high commercial value (up to $1,697 kg-1), they have a high exposure to international trade (McClenachan et al. 2016). Throughout the Pacific, the density of reef sharks has declined to 3–10% of pre-human levels (Nadon et al. 2012) and even the most well-managed Marine Protected Areas (MPAs) appear inadequate in maintaining healthy shark populations (Davidson & Dulvy 2017; Juhel et al. 2017).

It is unclear, however, if reported levels of dark diversity of sharks are due to local extirpations, or to a failure to detect remaining animals. Similar to most terrestrial vertebrates, sharks exhibit learning abilities linked to avoidance behaviour, and repeated

exposure to negative anthropogenic interactions may increase their elusiveness (Mourier et al. 2017). This raises the possibility that sharks' prevalence in marine habitats, even close to humans, may be greater than previously thought, with individuals being less detectable, therefore over-inflating the apparent level of dark diversity.

Environmental DNA (eDNA) analysis is based on the retrieval of genetic material naturally released by organisms in their environments and is emerging as a non-invasive method to detect and identify even rare and elusive species in a wide range of ecosystems (Bohmann et al. 2014), including marine waters (Miya et al. 2015; Thomsen et al. 2016). Here we assessed the potential of eDNA metabarcoding in providing a more accurate estimate of the dark diversity of sharks on coral reefs of the New Caledonian archipelago, by contrasting eDNA analysis with traditional Underwater Visual Censuses (UVC) and Baited Remote Underwater Video Stations (BRUVS) survey methods.

3.3 Results & Discussion

3.3.1 Lower dark diversity than previously estimated

Out of 26 historically present species in the regional pool (Tirard 2011), only nine species of sharks were detected in 2,758 UVCs and 385 BRUVS (Fig. 3.1A-C). The dark diversity of sharks was thus initially estimated at 65% of the regional pool (i.e. 17 species were not detected) using traditional survey methods. Despite two orders of magnitude less sampling effort, eDNA detected 44% more species than UVC or BRUVS; with only 22 samples, 13 shark species were detected, reducing the previously estimated dark diversity to 50% of the regional pool (13 undetected species). Six species were only detected by eDNA, three species only by UVC and BRUVS (of which, one species only by BRUVS), while six species were detected by all three methods (Fig. 3.2).

Sharks were observed in only 15% of UVCs (N=405 out of 2,758) and 54% of BRUVS (N=207 out of 385) (Fig. 3.1D). Furthermore, shark diversity was low for each sample, with only 3% of UVCs and 23% of BRUVS recording more than one species. When excluding the two most common species, the grey reef shark (*Carcharhinus amblyrhynchos*) and the whitetip reef shark (*Triaenodon obesus*), other shark species were observed in only 2% of UVCs and 13% of BRUVS.

In contrast, sharks were detected in 91% of eDNA samples (Fig. 3.1D), with 64% of samples revealing at least two species. No sharks were detected in the negative controls (see Methods). Even after excluding grey reef and whitetip reef sharks, 68% of eDNA samples revealed one or more shark species. The mean shark diversity per sample was significantly different between techniques (Kruskal-Wallis, $p<0.001$), with eDNA detecting at least three times more species (2.5±1.9; Dunn's tests, $p<0.001$; Fig. 1E) than BRUVS (0.8±0.8) and UVCs (0.2±0.5). These results suggest that the level of dark diversity of sharks on New Caledonian coral reefs is much lower than previously estimated with traditional techniques.

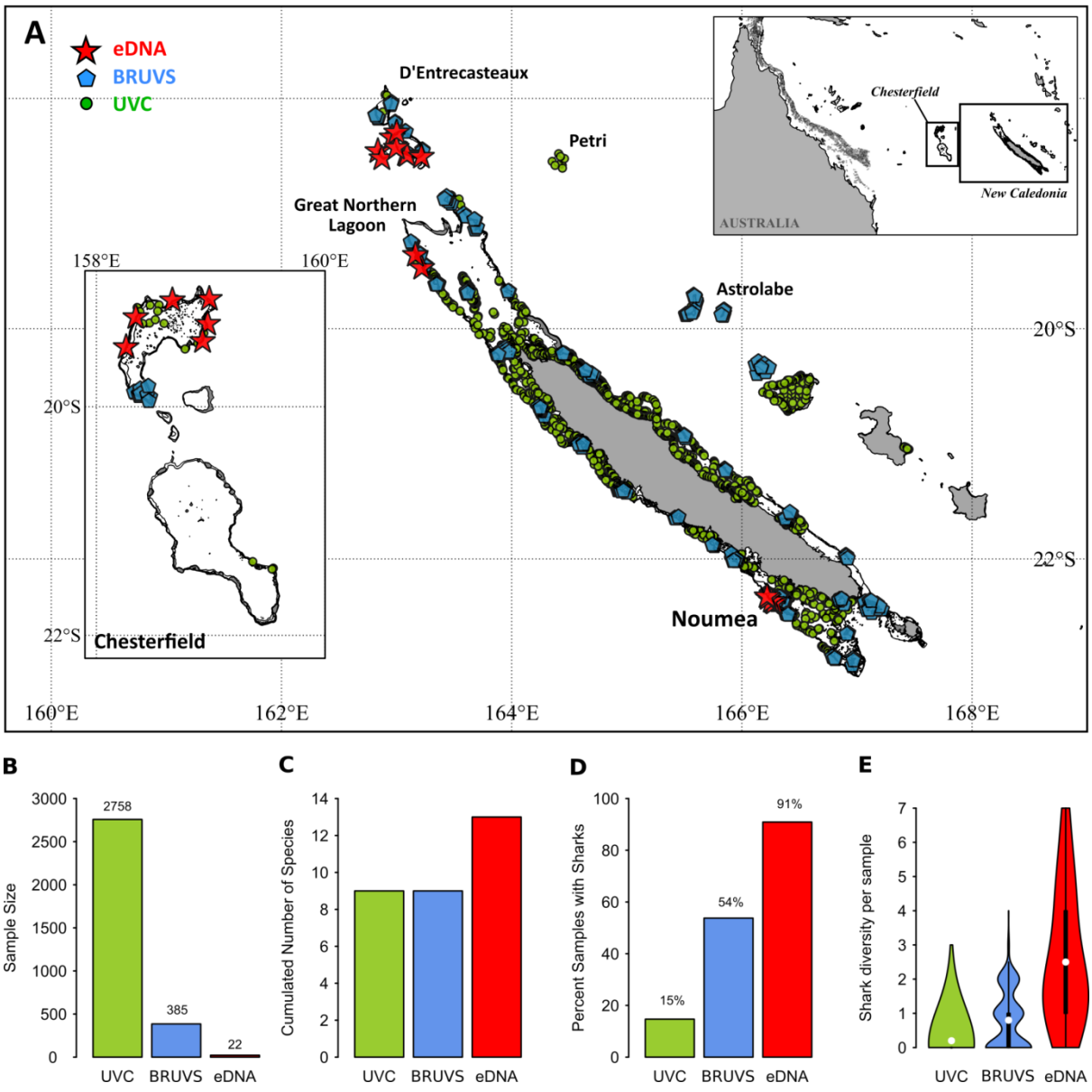


Figure 3.1 Sampling design and analyses of surveys across the New Caledonian archipelago, South-Western Pacific. (**A**) Sampling design in the New Caledonian archipelago (red stars = environmental DNA – eDNA, blue pentagons = Baited Remote Underwater Video Stations – BRUVS, and green dots = Underwater Visual Censuses – UVC. (**B**) Sample size. UVC: N=2,758. BRUVS: N=385. eDNA: N=22. (**C**) Cumulated number of shark species detected. (**D**) Frequency of samples with sharks detected. (**E**) Violin plot showing detected shark species richness, significantly different between techniques (Kruskal-Wallis, $p<0.001$), with eDNA detecting more shark species (2.5±1.9) compared to BRUVS (0.8±0.8) and UVC (0.2±0.5) (Dunn's tests, $p<0.001$). White dots are mean values; thick black bars correspond to interquartile ranges; thin black lines are 95% confidence intervals.

For any given time-interval, environmental DNA detects biodiversity at a greater spatio-temporal scale than the traditional methods used in this study; this may de facto increase the number of species detected by eDNA analysis (Goldberg et al. 2016). For example, UVCs sample limited visual areas (<500m2), within specific habitats, over short temporal periods (<2h) (D'agata et al. 2016). Similarly, although the bait plume from BRUVS can attract sharks from surrounding habitats, their detection capabilities are constrained by limited visual range (<50m) and operation time (a few hours) (Juhel et al. 2017). Contrastingly, eDNA may detect species at a greater temporal scale (a few hours to a few days) due to the persistence of cellular material in the water (Minamoto et al. 2017; Thomsen, Kielgast, Lars L. Iversen, et al. 2012). Moreover, as water masses are in constant movement (particularly along the outer slopes of coral reefs), eDNA transported from different habitats (e.g. open ocean) could potentially result in an overestimation of species richness in a given habitat. However, apart from occasional coral reef transients (e.g. the great hammerhead, *Sphyrna mokarran*), all shark species detected by eDNA are regularly observed in coral reefs habitats (Compagno et al. 2005) (Fig. 3.2). As such, it is unlikely that our diversity estimate has been inflated by oceanic inputs of external eDNA. Moreover, recent studies have indicated that eDNA analysis is powerful enough to distinguish species assemblages separated by small distances, even when comparing inshore and offshore habitats (Port et al. 2016; O'Donnell et al. 2017).

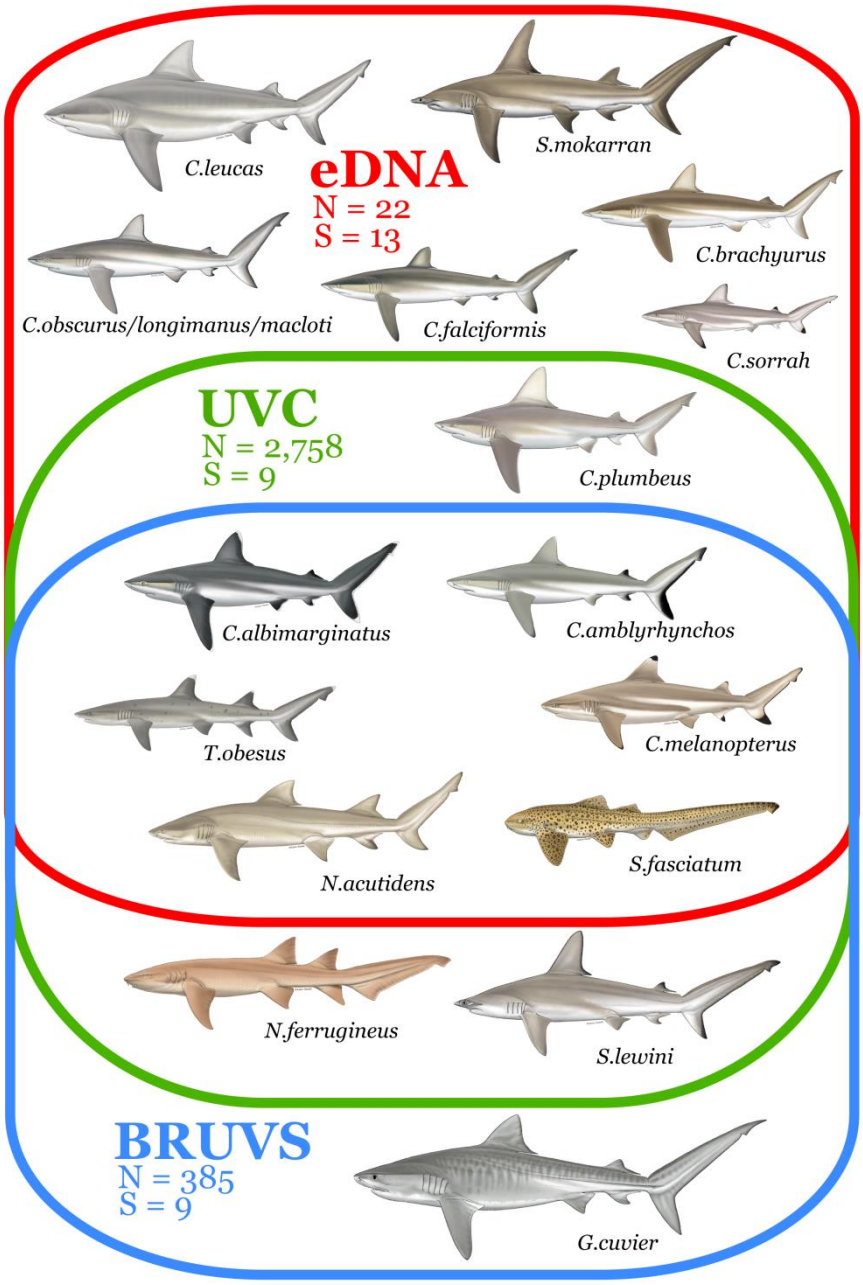


Figure 3.2 Detection of shark species among methods. Venn diagram showing the species detected by environmental DNA – eDNA (N=22 samples, S=13 species), Underwater Visual Censuses – UVC (N=2,758 samples, S=9 species) and Baited Remote Underwater Video Stations – BRUVS (N=385 samples, S=9 species). Scientific drawings courtesy of Marc Dando.

3.3.2 Persisting shark populations in human-impacted areas

Scientific literature has repeatedly highlighted the footprint of anthropogenic activities on shark populations worldwide (e.g. (Edgar et al. 2014; Robbins et al. 2006)). Similarly, BRUVS and UVC surveys suggest that sharks are quasi-absent near Nouméa, the capital city of New Caledonia (D'agata et al. 2016). Human-induced behavioural changes have been observed in both terrestrial (Frid & Dill 2002) and marine vertebrates (French et al. 2011). Behavioural changes in fish, due to differential responses to diver presence, may result in an overestimation of fish densities in marine reserves, and an underestimation in impacted areas, (Goetze et al. 2017). Little is known about the extent of similar behavioural sampling bias in sharks (Mourier et al. 2017). However, our results suggest that this bias may play a role in shark detection, particularly near densely populated areas, where eDNA detected a significantly greater diversity of sharks compared to UVCs and BRUVS (Fig. 3.3A, Kruskal-Wallis test, $p<0.001$, Dunn's tests, $p<0.001$). Failure of traditional methods to detect comparative levels of shark diversity around human-populated areas may be a reflection of more than just low shark densities, but also of avoidance behaviour in remaining individuals. Conversely, as sharks in relatively undisturbed areas may display curiosity or naivety (Goetze et al. 2017), we would expect BRUVS and UVCs to reveal high shark diversity in 'wilderness' areas located over 20 hours travel time from the main regional city (D'agata et al. 2016; Maire et al. 2016). However, eDNA detected three times more species in these areas (3.1±2.0) than BRUVS (1.3±0.8) and UVCs (0.9±0.8; Fig. 3.3B; Kruskal-Wallis test, $p<0.001$, Dunn's tests, $p=0.001$), demonstrating that eDNA appears more effective at estimating dark diversity, even when animal behaviour may bias direct observations positively or negatively. Due to spatial heterogeneity of our sampling design between the three techniques (Fig. 3.1A), we performed the same analysis for overlapping collection sites, and found very consistent results (Supplementary material 3.1).

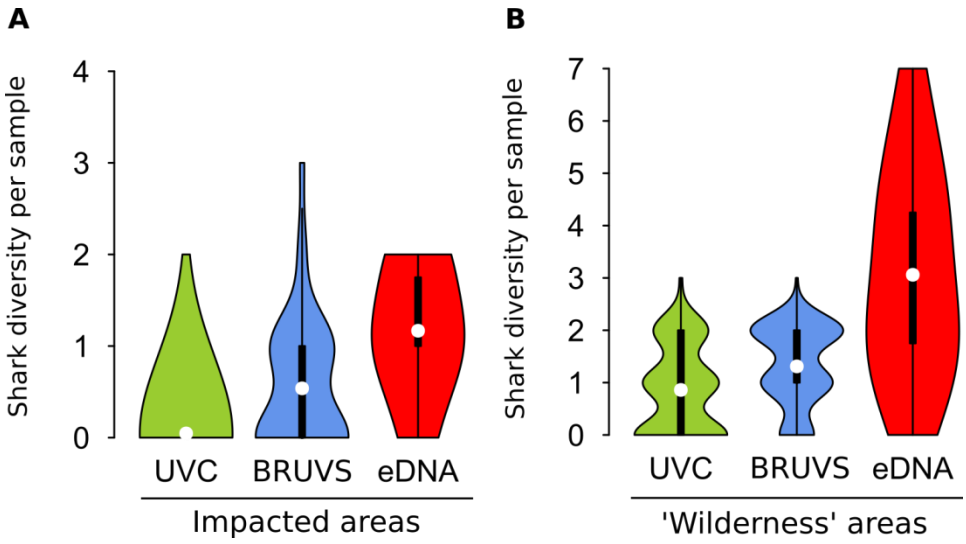


Figure 3.3 Number of shark species per sample in contrasts of human impacts. Violin plot showing detected shark species richness by the different methods in (**A**) impacted areas (Nouméa, the capital city), and (**B**) 'wilderness' areas (Chesterfield, D'Entrecasteaux, Great Northern Lagoon, Petri and Astrolabe). White dots are mean values; thick black bars correspond to interquartile ranges; thin black lines are 95% confidence intervals. Differences between methods are highly significant for both types of areas (Kruskal-Wallis tests, $p<0.001$), with eDNA detecting more species per sample than BRUVS and UVC (Dunn's tests, $p<0.001$).

3.3.3 **Increased species detectability revealed by rarefaction curves**

Regional species diversity may be assessed by rarefaction curves linking the number of detected species to sampling effort (Gotelli & Colwell 2001). Rarefaction curves associated with the three methods (UVC, BRUVS and eDNA) were fitted using six models with contrasted features (asymptotic vs. non-asymptotic; two vs. three parameters). The best model was non-asymptotic (power) for UVC-based survey, and asymptotic for BRUVS and eDNA surveys (negative exponential and rational function respectively) (Table 3.1). The rarefaction curve for BRUVS reaches nine species for the New Caledonian archipelago after 385 samples (Fig. 3.4B-C). We then fitted the same asymptotic model (rational function) to the three rarefaction curves to compare their asymptotes. We show that doubling BRUVS sampling effort would result in the detection of only a single additional species (Table 3.1 & Fig. 3.4D). Meanwhile, the rarefaction curve for UVC attains nine species after 2,758 UVCs (Fig. 3.4B-C). According to the common model, doubling the number of UVCs would be required to detect one additional shark species (Table 3.1 & Fig. 3.4D), and both BRUVS and UVCs rarefaction curves plateau at ten species. Contrastingly, a few hundred eDNA samples

could provide an accurate assessment of regional shark diversity (Fig. 3.4B-D, Table 3.1), requiring much less time and equipment than traditional survey methods, and rapidly revealing a considerable proportion of unseen shark species by classical methods that tend to overlook rare and elusive species in regional inventories (Fig. 3.4A). In addition, eDNA sampling was only conducted during three months in 2015, while BRUVS were deployed over a three-year period (2012 to 2014) and UVC surveys were conducted between 1986 and 2014, reinforcing the potential of eDNA metabarcoding. Indeed, our results are very conservative in the sense that traditional methods were carried out over a large spatio-temporal scale, thus more likely to sample unique events like rare species appearance or migration.

Table 3.1 Models fitted for species rarefaction curves obtained from Underwater Visual Census (UVC), Baited Remote Underwater Video Station (BRUVS) and environmental DNA (eDNA), using the nls function in the stats package and the Akaike Information Criterion corrected for small sample bias (AICc, package AICmodavg). Best fitting models are in bold characters and the overall best fitting common model is in red.

Model	**Formula**	**Asymptotic**	**Number of parameters**	**AICc**		
				UVC	**BRUVS**	**eDNA**
Power	$S = aX^b$	No	2	**-4345**	528	14
Exponential	$S = a + b\log(X)$	No	2	-191	210	23
Negative exponential	$S = a(1 - e^{-bX})$	Yes	2	4940	281	24
Negative exponential	$S = a + (b - a)e^{-cX}$	Yes	3	-609	**-675**	-19
Monod	$S = a/(1 + bX^{-1})$	Yes	2	2321	-185	-5
Rational function	$S = (a + bX)/(1 + cX)$	Yes	3	-2562	-421	**-49**

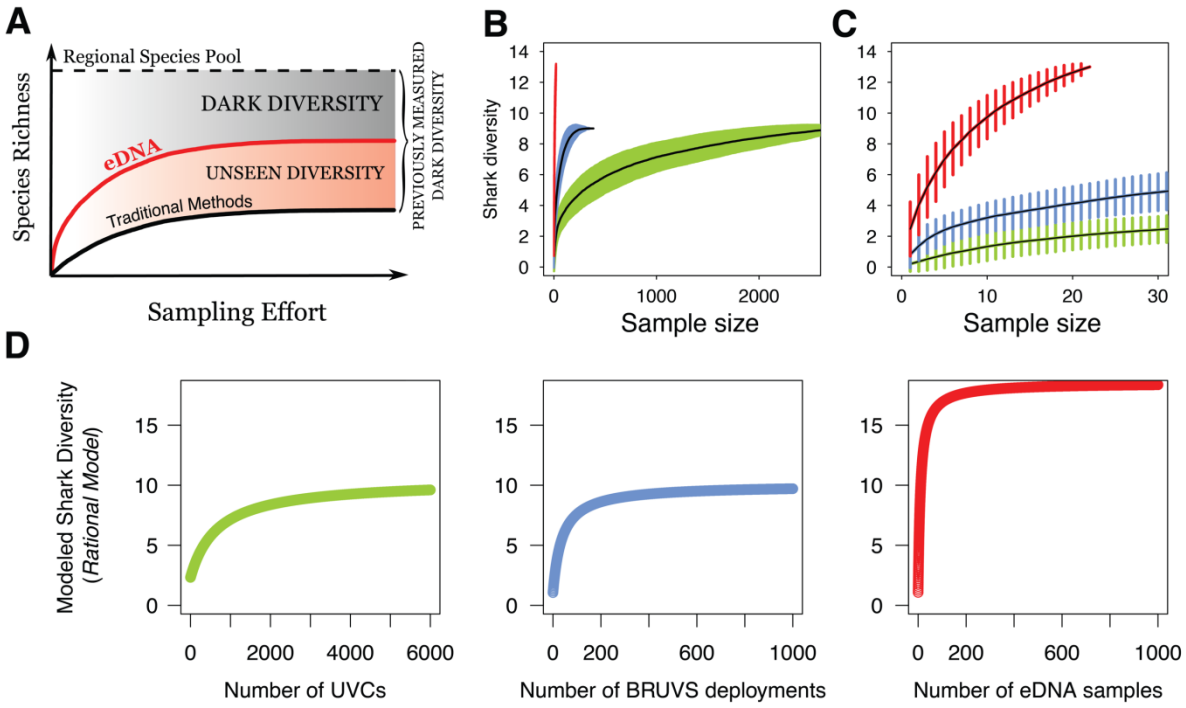


Figure 3.4 Sample-based rarefaction curves. (**A**) Theoretical illustration of dark diversity measured by traditional methods, simply unseen but illuminated by eDNA, revealing a lower amount of dark diversity (absent species). Rarefaction curves showing accumulated sampled shark diversity measured by the different techniques (green = Underwater Visual Census – UVC, blue = Baited Remote Underwater Video Stations – BRUVS, and red = environmental DNA – eDNA): (**B**) based on all samples, (**C**) zoomed in to 30 samples. Error bars indicate standard deviation. (**D**) Estimated rarefaction curves for UVC, BRUVS and eDNA when increasing sampling effort, based on the best common model (rational function).

3.3.4 **Limits and uncertainty of species detection**

Using the 127-bp COI fragment, we did not find an unequivocal correspondence between Molecular Operational Taxonomic Units (MOTUs) and species, as some MOTUs had 100% sequence identity matches with multiple species in the BOLD database (Bakker et al. 2017). This issue mainly concerns the Carcharhinus genus, recognized to be taxonomically problematic and polyphyletic (Sorenson et al. 2014). Therefore, although 16 different shark MOTUs were identified, we opted for the conservative approach of merging MOTUs to present the minimum species richness (13 species). With a trade-off between primer universality and taxonomic resolution (Thomsen et al. 2016), the imperfect nature of currently available metabarcoding primers introduces a degree of uncertainty regarding the identification of certain species (Stat et al. 2017).

Only three species (tiger, scalloped hammerhead and nurse sharks) that were observed by UVC and/or BRUVS, were not detected by eDNA (Fig. 3.2). The primers used in this study

have already been shown to be capable of amplifying eDNA from both Sphyrnidae (hammerhead sharks) and *Galeocerdo cuvier* (tiger shark), but not Ginglymostomatidae (nurse sharks), explaining the absence of *Nebrius ferrugineus* (tawny nurse shark) from eDNA detections (Bakker et al. 2017). Previous studies have indicated that hammerhead and tiger sharks only occur at very low densities in New Caledonia (Juhel et al. 2017). Even though our results show the power and potential of eDNA for the detection of shark diversity, our sampling effort was insufficient for an exhaustive survey, due to stochasticity in PCR and sequencing, but also because these species may not have been present at the specific location and time of sampling. It is thus very likely that increasing eDNA sampling effort would detect them in New Caledonia, as the rarefaction curves show that eDNA can outperform the other methods in terms of species detectability (Juhel et al. 2017), but additional work on refining the design of primers is needed to cement eDNA as a standard tool for the study of dark diversity of sharks.

There are also important caveats associated with eDNA detection and traditional survey methods still have a number of advantages over eDNA methodologies. Foremost, eDNA-based methods cannot provide information on size, condition, developmental stage (eggs, larvae, juveniles, adults), sex, behavior and movement patterns of the target organism. Furthermore, when using typically maternally inherited mitochondrial markers, it remains impossible to distinguish hybrids, which may be the result of breeding between native and invasive species, from their maternal species. Finally, inferring species abundance from eDNA still remains a challenging but promising avenue and is a key area for further research (Bakker et al. 2017; Jo et al. 2017; Yamamoto et al. 2016).

3.3.5 **New light for megafauna conservation**

In this study, we highlight the potential of eDNA metabarcoding for the rapid assessment of elusive megafauna species richness, and for the first time, for the determination of the extent of purported local species extirpations. We call for the introduction of eDNA assessments to complement traditional survey methods for the improvement of species detection, and hence, more efficient conservation strategies for threatened and elusive megafauna.

Firstly, environmental DNA allows for the reappraisal of previous estimates of species occurrences, which are used to define Criteria B of IUCN Red List Categories. With 46.8% of shark species data deficient in IUCN assessments (Dulvy et al. 2014), this

knowledge may markedly aid Red List classifications, and the design and implementation of future conservation measures. For instance, a species is classified as Vulnerable if its extent of occurrence, i.e. its continuous geographic range, is less than 20,000 km2 and declining. This kind of assessment is hardly achievable, trustable and repeatable with traditional sampling methods owing to low detectability. However, eDNA provides an affordable, powerful and standardized tool to assess large-scale occurrences, even for elusive megafauna. IUCN classification can also be based on population size (Criteria C) and, while much more research is needed in this field, quantitative methods to infer eDNA concentration offer promise towards good estimates of relative abundance (Yamamoto et al. 2016; Lacoursière-Roussel et al. 2016). Alternative approaches include detecting and distinguishing individuals of a given population as a measure of abundance. Differentiating populations from a single shark species has recently been achieved using eDNA (Sigsgaard et al. 2016). This strategy, even in its infancy, opens a new era in the field of population genetics using eDNA and in population size assessment potentially fuelling IUCN Criteria C. Monitoring shark populations of New Caledonia may greatly benefit from such methodological advances since the effectiveness of conservation efforts is still under scrutiny for megafauna (Juhel et al. 2017; Bauer et al. 2015).

Additionally, increased knowledge of dark diversity may guide the direction of conservation-based decision-making (Lewis et al. 2017). To halt biodiversity loss, it is imperative to understand why some species are missing from areas where biogeographic history, as well as current ecological and environmental conditions, predict their presence. Seeking for common characteristics (e.g. ecological needs, dispersal ability, body size) among species constituting the dark diversity can help identifying key determinants of vulnerability, decline, or extirpation, and guide appropriate management strategies (Moeslund et al. 2017). High proportions of dark diversity in a given region or area indicate the need for widespread conservation efforts across multiple species, while low proportions of dark diversity suggest that more tailored solutions are required to reduce pressures on specific species apparently missing. Areas showing high observed diversity and low dark diversity can be considered as refugia, and thus potential sources for recolonization. Consequently, such areas deserve high conservation priority (Lewis et al. 2017; Pärtel et al. 2013). By contrast, areas with relatively high dark diversity compared to the observed species richness need restoration efforts focusing on mitigation of species threats and increasing connectivity. In New Caledonia, mapping and monitoring the dark related to the observed

diversity of sharks along human gradients would provide a relevant indicator of wildlife status to inform management.

Large-scale efforts to restore local species pools are more feasible in terrestrial than in marine environments. For example, creating habitat connectivity to promote species re-establishment has previously been shown to be successful (e.g. (Di Minin et al. 2013)). However, this requires alleviation of the original stressors (through, for instance, protected areas) to avoid impacting newly reconnected populations (Di Minin et al. 2013). Assisted reintroduction is another potential solution in terrestrial habitats, albeit costly and controversial (Seddon et al. 2014; Weise et al. 2014). As there are less manipulative solutions for the marine environment, the discovery of remaining individuals close to human-impacted areas requires more immediate, alternative actions, such as the establishment of marine reserve networks, connecting suitable habitats, in order to preserve the remainder of the species and increase population densities by decreasing threats. Environmental DNA will most likely prove progressively useful in marine conservation, and hence, will be playing an increasingly important role in the formulation of policies to aid species conservation.

3.4 Materials & Methods

3.4.1 **Study sites.**

The New Caledonian archipelago is located in the south-western Pacific Ocean (Fig.3.1A). It is comprised of 'Grande Terre' surrounded by one of the largest barrier reefs in the world, and numerous isolated islands and remote reefs. Sampling occurred across the New Caledonian Archipelago, including waters from the Coral Sea Marine Park. Study areas encompassed a gradient of human density, from high population density (near the capital, Nouméa), to wilderness reefs at >20h travel time from the main regional city (Chesterfield Reefs, D'Entrecasteaux Reefs, Astrolabe Reefs, Petri, Great Northern Lagoon (Juhel et al. 2017; D'agata et al. 2016; Maire et al. 2016)). The regional pool of sharks in New Caledonia is inventoried at 49 species, including 26 shallow water species (Tirard 2011).

3.4.2 **Underwater Visual Census and Baited Remote Underwater Video Station datasets**

Underwater Visual Census (UVC) and Baited Remote Underwater Video Stations (BRUVS) protocols used in this study are described in detail in (D'agata et al. 2016) and (Juhel et al. 2017). Here, 2,758 UVCs were conducted by day in various coral reefs habitats at 1-15 m depth from 1986 to 2014 (Fig.3.1A). 385 BRUVS (Harvey et al. 2007) were deployed by day at a mean depth of 16 m (± 10 m SD; range = 3-48 m), in different coral reef habitats between September 2012 and October 2014. BRUVS are video systems that record for one hour in the presence of standardized bait (1 kg pilchards). Shark occurrence was measured through video analysis, and species identification was double-checked by trained operators.

3.4.3 **eDNA collection and sample processing.**

Environmental DNA samples were collected during September-November of 2015, on coral reef external slopes, with a reef topography ranging from 20 – 40 m. Each four-litre water sample consisted of two litres sampled at 5 m depth and two litres at 20 m depth, collected with a Niskin water sampler. After collection, water samples were individually covered and stored on ice prior to filtration. Water was subsequently filtered using sterile mixed cellulose esters (MCE) filters (Merck Millipore; 47mm diameter; 0.45 μm pore size), then stored at -20°C in 2.0 ml screw-cap microcentrifuge tubes containing silica beads, drying out the filters and preventing DNA degradation (Bakker et al. 2017). DNA was extracted from the filters using the DNeasy PowerSoil DNA Isolation Kit (Qiagen), following the manufacturer's protocol. Purified extracts were assessed for DNA concentration in a Qubit fluorometer (Thermo Fisher Scientific).

Strict adherence to contamination control was followed at all field and laboratory stages in order to prevent the occurrence of contamination, including the use of disposable gloves and single use-sterile collection bottles and filtration equipment, and the bleaching (50% bleach) of sampling devices and laboratory equipment and surfaces. Additionally, a dedicated controlled eDNA lab at the University of Salford, with separate rooms designated for the physical separation of eDNA extraction, pre-PCR preparations and post-PCR procedures, was used for all laboratory work. To identify any potential contamination,

negative-control DNA extraction blanks (elution buffer from extraction kit) and PCR blanks were also ran (Bakker et al. 2017).

3.4.4 **Library preparation and sequencing**

An elasmobranch-specific cytochrome c oxidase I (COI) primer set was used for the amplification of eDNA metabarcoding markers. The previously published primer set consisted of a novel reverse primer 'Shark COI-MINIR' 5'-AAGATTACAAAAGCGTGGGC-3' (46) and two universal fish barcoding forward primers FishF2 5'-TCGACTAATCATAAAGATATCGGCAC-3' and FishF1 5'-TCAACCAACCACAAAGACATTGGCAC-3' (Fields et al. 2015; Ward et al. 2005), yielding an amplicon of 127 bp. Samples were sequenced in a single multiplexed Illumina MiSeq run, along with samples from a related project, which are not included in this study, for a total of 96 samples including two negative controls, using four sets of 24 primers with attached 8-base sample-specific oligo-tags differing in at least three bases (Supplementary material 3.2) (Guardiola et al. 2015). To increase variability of the amplicon sequences, a variable number (two, three or four) of fully degenerate positions (Ns) was added at the beginning of each primer (Wangensteen & Turon 2017). The full, sequenced PCR product, consisted of 195 bp, including the amplicon, primers, sample tags and leading Ns.

For PCR amplification, a single step protocol was used, directly attaching the 8-base tagged primers. The PCR mix recipe was as follows: a total volume of 20 µl included 2 µl 10x buffer (BioLine), 0.6 µl 50 mM MgCl (BioLine), 0.5 µl of each of the 5 µM forward primers (Eurofins), 1 µl of the 5 µM reverse primer, 0.2 µl 10 mM dNTP mix (BioLine), 0.2 µl BioTaq DNA polymerase (5 u/µl, BioLine), a standardized amount (10 ng) of the filter-extracted eDNA template, and 13 µl sterile water. The PCR profile included an initial denaturing step of 95ºC for 15 min, 35 cycles of 94ºC 1 min, 52ºC 1 min and 72ºC 1 min and a final extension step of 72ºC for five minutes. The quality of all amplifications was assessed by electrophoresis, running the products through a 1.5% agarose gel stained with GelRed (Cambridge Bioscience) and visualized on a UV light platform. All PCR products (including two PCR negative controls, were pooled by marker into four multiplex sample pools and purified using MinElute columns (Qiagen). Four Illumina libraries were subsequently built from the four pools, using the NextFlex PCR-free library preparation kit (BIOO Scientific). The libraries were quantified using the NEBNext qPCR quantification kit (New England

Biolabs) and pooled in equimolar concentrations along with 1% PhiX (v3, Illumina) serving as a positive sequencing quality control. The libraries with a final molarity of 8 pM were sequenced on an Illumina MiSeq platform in a single MiSeq flow cell using v2 chemistry (2 x 150 bp paired-ends).

3.4.5 **Bioinformatics analyses**

The bioinformatic analysis was based on the OBITools metabarcoding software suite (Boyer et al. 2016). The pipeline used for data analysis is summarized in Supplementary material 3.3. Quality of the reads was assessed using FastQC. Paired-end reads were aligned using illuminapairedend, and alignments with quality scores >40 were kept. The aligned dataset was demultiplexed using ngsfilter. A length filter (obigrep) was applied to the aligned reads (120-135 bp) and reads containing ambiguous bases were removed. The reads were then dereplicated using obiuniq and a chimera removal step was performed using the uchime-denovo algorithm (Edgar et al. 2011) implemented in vsearch (Rognes et al. 2016). The Molecular Operational Taxonomic Units (MOTUs) were delimited using the sumaclust algorithm (Boyer et al. 2016) with a constant similarity threshold of 99%. Taxonomic assignment of the representative sequences for each MOTU was performed using the ecotag algorithm (Boyer et al. 2016). We built a bespoke elasmobranch reference database using a custom R script for retrieving all COI elasmobranch sequences available from the BOLD database (Ratnasingham & Hebert 2007), and subsequently selecting those that included our 127 bp target fragment. The custom R script is available from http://github.com/metabarpark/R_scripts_querying_databases. In order to add homologous sequences from other, non-elasmobranch taxa, an *in silico* PCR was performed against release R117 of the EMBL-EBI database using ecoPCR (Ficetola et al. 2010). Subsequently, the obtained reference sequences were added to the elasmobranch sequences obtained from BOLD. These additional reference sequences were added to our elasmobranch database to avoid the incorrect assignment of amplified sequences, belonging to other taxa, to elasmobranchs. This combined reference database is available from http://github.com/metabarpark/Reference-databases. The final refining of the dataset included taxonomy clustering of MOTUs assigned to the same species.

The risk of contamination adds to the challenges associated with eDNA metabarcoding (Goldberg et al. 2016; Thomsen & Willerslev 2015), with the possibility of

introducing false positive results. It is likely to detect a species represented by a single sequence read in a sample, but the possibility of contamination or sequencing error cannot be excluded as the potential cause of MOTU detection. We subsequently adopted a conservative approach to our analyses and removed single read MOTUs from our samples to avoid potential false positives.

3.4.6 **Statistical analyses**

Given the violation of the normality assumption and the unbalanced design with different number of samples depending on the techniques, Kruskal-Wallis tests followed by Dunn's tests were performed to test for differences in shark diversity per sample among techniques. The vegan package was used for rarefaction analyses followed by model fitting using the nls function in the stats package and the Akaike Information Criterion corrected for small sample bias (AICc, package AICmodavg). Models were fitted for the three methods independently (UVC, BRUVS and eDNA, Table 3.1), subsequently, a common model was selected by comparing AICc for the three methods simultaneously. Statistical analyses were performed in the R program environment (R Development Core Team 2012, version 3.4.0).

3.5 Acknowledgements

We thank the crew of the research vessels Amborella (Government of New Caledonia), Alis (IRD) and Plan B (Waitt Institute) for their help with data collection during oceanographic campaigns. We thank Gerard Mou-Tham and many others for UVC collection and BRUVS analyses. We also would like to thank Marc Dando for the scientific drawings used in the manuscript. Funding: This work is an outcome of the Pristine Reefs, APEX and e(lasmo)DNA projects co-funded by the Total Foundation, the Pew Charitable Trusts, the Government of New Caledonia, and the University of Salford R&E strategy funding

4 Chapter IV

Tropical biodiversity assessment of shelf eukaryotic communities via eDNA metabarcoding

Judith Bakker, Owen S. Wangensteen, Charles Baillie, Dayne Buddo, Demian D. Chapman, Austin Gallagher, Tristan L. Guttridge, Heidi Hertler, & Stefano Mariani

Author contributions: J.B. and S.M. conceived, designed and coordinated the study; J.B., S.M., D.D.C., A.G., D.B., T.L.G., and H.H. contributed to fieldwork and sample collection; Laboratory experiments and data analyses were conducted by J.B., O.S.W. and C.B.; J.B. wrote the manuscript; all authors read and commented on the manuscript.

4.1 Abstract

The understanding of marine communities and their functions in an ecosystem relies heavily on our ability to detect and monitor species distributions and abundances. However, one of the most critical issues in marine conservation is the lack of efficient and reliable tools to comprehensively assess and quantify biodiversity. Currently, the use of environmental DNA (eDNA) metabarcoding is increasingly being applied for the rapid assessment and monitoring of aquatic species. But this has thus far mostly pertained to the identification of a few specific taxa or groups simultaneously. However, the application of eDNA on a much broader taxonomic scale, such as for the description of complete marine eukaryotic communities, may be greatly beneficial for a more holistic perception of biodiversity and ecosystem functioning. Here we investigate the potential of eDNA COI metabarcoding, for the characterisation of the biodiversity of complex natural marine communities in tropical coastal shelf habitats. We screened 67 samples from five Caribbean locations and detected a high level of species richness. However, a disproportionally large number of eukaryote taxa remained unassigned, suggesting that the sampled communities host an astonishing amount of yet undescribed micro-eukaryotic diversity. Nonetheless, it was possible to characterize spatial patters of diversity between and within the five locations.

4.2 Introduction

The foundation for ecosystem research and the effective management of biodiversity in natural systems is the knowledge of what species are present in that ecosystem (Mace et al. 2012). And while rapid assessment and monitoring of biodiversity are imperative, the time and resources required to generate the necessary data are a major constraint in ecological research and conservation management. Marine habitats cover >70% of the Earth's surface and constitute one of the vastest and richest biomes. Most recent estimates, all based on indirect approaches, suggest that there are millions of marine eukaryotic species (Appeltans et al. 2012; Costello et al. 2012; Mora et al. 2011), with the vast majority being small (< 1mm), cryptic and currently unknown to science (Leray & Knowlton 2016). Accordingly, there is a pressing need to measure marine biodiversity and to quantify the rate at which it is changing. However, many traditional survey methods are inefficient, selective, potentially destructive, and dependent on a declining taxonomical expertise (Wheeler 2004), indicating the need for alternative methods for biodiversity assessment and monitoring of aquatic species.

Molecular based methods are revolutionizing the analysis of biodiversity, as they offer advantages over traditional visual morphological survey methods. Recent developments in metabarcoding, coupled with high-throughput sequencing, have made it possible to detect thousands of species present in both bulk samples, and in samples of environmental DNA (eDNA), such as soil, sediment, air, and water. Environmental DNA metabarcoding utilizes universal PCR primers to mass-amplify informative, orthologous gene regions (DNA barcode sequences) from complex samples. These are subsequently sequenced and identified using molecular reference databases (Hajibabaei et al. 2011; Taberlet et al. 2012). Metabarcoding approaches have been successfully employed to characterize specific marine plankton communities in natural seawater samples, such as zooplankton, mesozooplankton and full eukaryotic plankton diversity (Djurhuus et al. 2018; Deagle et al. 2017; Chain et al. 2016; López-escardó et al. 2018; de Vargas et al. 2015; Villarino et al. 2018). Marine benthic communities have also been characterized, including communities from soft sediments, containing mostly small-sized organisms, (Guardiola et al. 2015; Lejzerowicz et al. 2015; Pawlowski et al. 2014), and hard-bottoms (Leray and Knowlton, 2015; Wangensteen, Palacín, Guardiola, 2017), featuring complex eukaryotic communities with body sizes spanning several orders of magnitude.

There are many potential benefits that eukaryotic, whole-community metabarcoding of marine eDNA could bring to biodiversity assessment and monitoring, such as using direct measurements of biodiversity, instead of relying on biodiversity indicators (Aylagas et al. 2016; Djurhuus et al. 2017; Lindenmayer & Likens 2011; Rees et al. 2014). Additionally, as it would allow for the detection of virtually all species without a priori knowledge of their presence in a water body, it can be used for the detection of 'hidden diversity' in species assemblages (Lindeque et al. 2013), and as such, as a powerful tool in the early detection of alien species (Zaiko et al. 2015), and of community structure changes in response to environmental disturbances (Bik et al. 2012; Bucklin et al. 2016). Moreover, the gains in cost-effectiveness, reproducibility, comprehensiveness, and the potential of multiple trophic levels being evaluated simultaneously, could make it easier to implement large-scale monitoring of biodiversity trends (Bourlat et al. 2013). Such approaches to community assessment could greatly assist efforts associated with environmental management, biodiversity conservation and policy-making (Baird et al. 2012; Chain et al. 2016; Goodwin et al. 2017).

Studies targeting marine eukaryotic community diversity have as of yet been very scarce (Djurhuus et al. 2017; Drummond et al. 2015; Kelly et al. 2016). And while the use of a single marker would add to the cost and resource-effectiveness of eukaryotic eDNA community metabarcoding, it requires a nearly-universal primer set, capable of amplifying the wide taxonomic array present in any eukaryotic community, while excluding the amplification of prokaryote taxa. Few recent studies have employed the COI (cytochrome oxidase I) marker region (Hebert, Ratnasingham, et al. 2003) for this purpose (Deagle et al. 2017; Leray & Knowlton 2017), but universal primers have usually been restricted to markers with more conserved regions such as the nuclear small-subunit 18S rRNA region (Guardiola et al. 2015) and the mitochondrial 16S region (Kelly et al. 2016). Metabarcoding studies of marine zooplankton have used various regions of both 18S and 28S, but due to their relatively conserved sequences, it is often impossible to distinguish taxa at the species, genus or even family level (Tang et al. 2012). Consequently, these markers underestimate species diversity, while species-level resolution may be essential for the description of community structure, the detection of rare or invasive species (Comtet et al. 2015; Aylagas et al. 2016), and the implementation of monitoring-based management and conservation measures. In particular because species level identification will provide more information towards the evaluation of the nature of potential community changes and the biological and/or environmental mechanisms responsible for these changes (Mackas & Beaugrand

2010). The identification of cold vs. warm water species for example, may help identify changes in local water temperatures.

Therefore, more variable markers, with broad taxonomic resolution, are required for community wide, high-resolution discrimination of species. The mitochondrial COI barcode region is one of the most commonly used regions for the analysis of species diversity among marine animals (Bucklin et al. 2011). While the use of COI as a metabarcoding marker has been criticized, arguing that the high rates of sequence variability impair the design of truly universal primers and hamper the bioinformatic analysis (Deagle et al. 2014), currently no other genetic region is represented in taxonomically verified databases with sequences covering the same number of taxa. Moreover, the high mutation rate of COI may ensure unequivocal identification at the species level across the vast majority of taxa. Although no gold standard universal metabarcoding primer set, for highly variable markers such as COI, that is able to amplify the full taxonomic range of a eukaryotic community has been identified (Coissac et al. 2012; Deagle et al. 2014; Riaz et al. 2011), it has been shown that the taxonomic coverage and resolution provided by degenerate COI primers (primer sets that have one or more degenerate positions incorporated in either one or both of the forward and reverse primers), make them valuable metabarcoding markers for biodiversity assessment (Elbrecht & Leese 2016; Clarke et al. 2017). Most recently, a degenerated version of the established COI internal primer set (Leray et al. 2013) amplifying a 313 bp region, has been described (Wangensteen, Palacín, Guardiola, 2017). This 'Leray-XT' primer set features a high number of degenerate positions, including 2 deoxyinosine nucleotides (a nucleotide that complements any of the four natural bases) in the fully degenerated sites of the sequence, enhancing universality in the amplification of the COI fragment in most eukaryotic groups. These primers have been shown to reveal greater biodiversity through increased resolving power at the species level, compared to 18S primers applied to the same samples (Wangensteen, Palacín, Guardiola, 2017).

Here, we evaluate the pertinence of the Leray-XT primer set, for the assessment of marine eukaryotic biodiversity in taxonomically complex water samples. We applied metabarcoding of total eDNA extracted from natural, unfractionated marine water samples, in an attempt to characterize the eukaryotic community, profile biodiversity and to assess spatial patterns among and between five different locations in the Caribbean basin.

4.3 Material & Methods

4.3.1 Water sampling

Field sampling was devised to test whether spatial patterns of eukaryotic biodiversity could be detected across five Caribbean locations (Figure 1). In Belize sampling took place around the partially submerged Glover's Reef atoll, which is part of the Mesoamerican Barrier Reef. In the Turks & Caicos Islands, samples were collected around South Caicos. In the Bahamas, the sampling was conducted around the islands of Bimini. In Jamaica the two main sampling areas where Montego Bay and Discovery Bay, and in the British Virgin Islands, samples were collected around Tortola, Virgin Gorda and Eustatia Island. Within each location, samples were collected from different types of habitats (See supplementary material 1 for sampling specifics). During February and March of 2015 (Jamaica, Belize, Turks & Caicos and The Bahamas), and February and March 2017 (British Virgin Islands), a total of 68 water samples, of 4 litres each, were collected with either a Kemmerer type water sampler or directly with a plastic collection bottle.

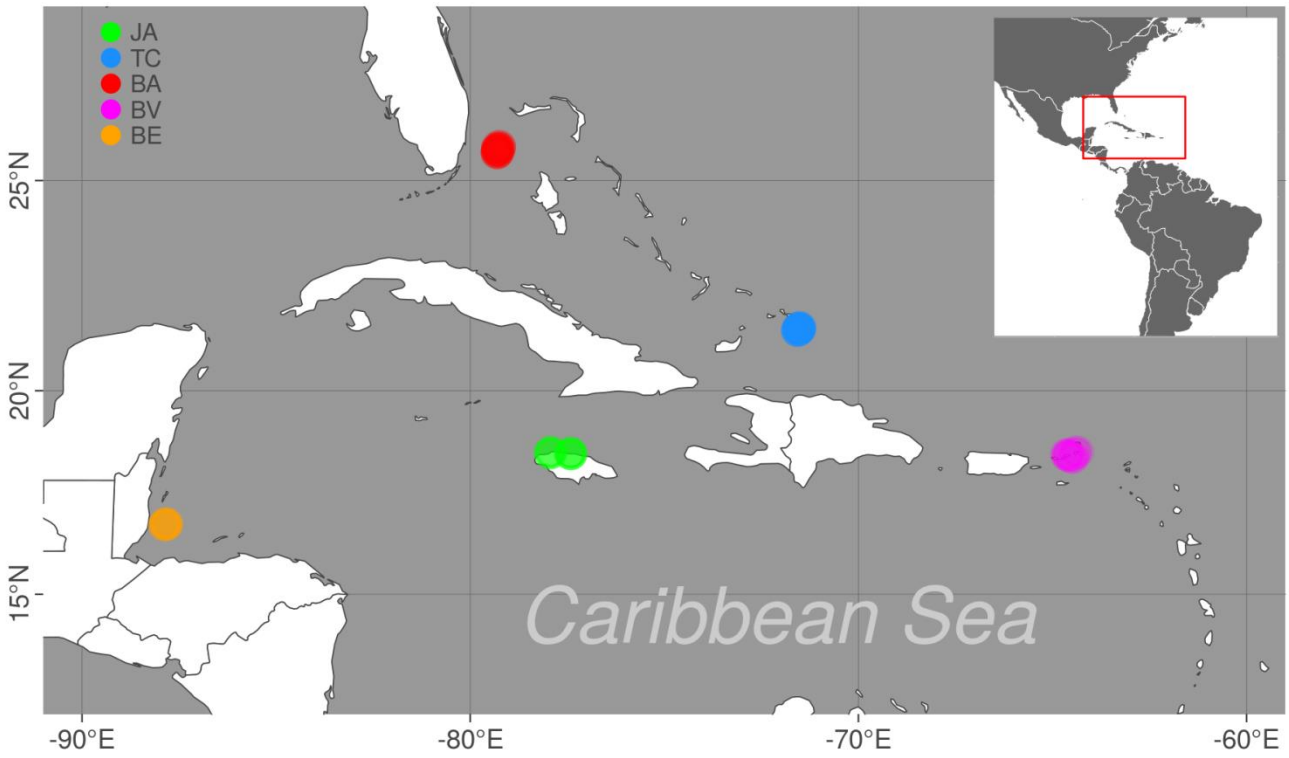


Figure 4.1 Map of Caribbean sampling locations. (**BA**) Bahamas, (**BE**) Belize, (**BV**) British Virgin Islands, (**JA**) Jamaica and (**TC**) Turks & Caicos.

4.3.2 **Sample processing and DNA extraction**

After collection, the water samples were individually covered and stored, in the dark and on ice, during transport to the local laboratory facilities. Vacuum filtration was carried out within two hours after collection. The sterile mixed cellulose esters (MCE) filters (Merck Millipore; 47 mm diameter; 0.45 µm pore size) containing sample filtrates were stored in 2.0 ml screw-cap microcentrifuge tubes containing silica beads. The silica beads function as a desiccator, drying out the filters and hence preventing the DNA from degrading (Bakker et al. 2017). The advantages of using silica beads instead of a liquid for DNA preservation, specifically high-grade ethanol, for DNA preservation are the prevention of leakages and the complications related to shipping/traveling with flammables. The sample filters were subsequently stored at -20°C until extraction. DNA was extracted from the filters with the DNeasy PowerSoil DNA Isolation Kit (Qiagen), following the manufacturers' protocol. Purified extracts were assessed for DNA concentration in a Qubit fluorometer (Thermo Fisher Scientific).

4.3.3 **Library preparation and sequencing**

For the 5’ region of COI, we used a novel, highly degenerated primer set (Wangensteen, Palacín, Guardiola 2017) yielding an amplicon of 313 bp. The reverse primer jgHCO2198 5'-TAIACYTCIGGRTGICCRAARAAYCA-3' (Geller et al. 2013), and a new forward primer miCOIintF-XT 5'-GGWACWRGWTGRACWITITAYCCYCC-3', which was modified from the miCOIintF primer (Leray et al. 2013) by incorporating two more wobble bases (equimolar mixtures of two or more different bases at a given position within the sequence) and two inosine nucleotides (capable of base-pairing with any of the four nucleotides) in most degenerate positions, in order to increase universality across eukaryotic groups (Wangensteen, Palacín, Guardiola 2017). For two multiplex Illumina sequencing runs we used primers with attached 8-base sample-specific oligo-tags differing in at least 3 bases (Guardiola et al. 2015). In order to increase variability of the amplicon sequences, a variable number (2, 3 or 4) of fully degenerate positions (Ns) was added at the beginning of each primer (Wangensteen & Turon 2017). The full, sequenced PCR product, consisted then of 389 bp, including the amplicon, primers, sample tags and leading Ns. For PCR amplification, a two-step protocol was used, attaching the 8-base tagged primers, after an initial amplification, in the second PCR. The mix recipe for the first PCR included 10 µl AmpliTaq Gold DNA polymerase, 1 µl of each 5 µM forward and reverse primers, 0.16 µl bovine serum albumin (BSA), 5.84 µl sterile water and a standardised amount (10 ng) of the filter-extracted eDNA template, in a total volume of 20 µl per sample. The recipe for the second PCR was identical, except that now the forward and reverse 8-base tagged primers were used. The PCR profile included an initial denaturing step of 95 ºC for 10 min. 35 cycles of 94 ºC 1 min, 45 ºC 1 min and 72 ºC 1 min and a final extension step of 5 min at 72 ºC. The profile for the second stage PCR was identical. The quality of all amplifications was assessed by electrophoresis, running the products through a 1.5% agarose gel stained with Gel Red (Cambridge Bioscience) and visualized on a UV light platform. All second stage PCR products, including two extraction and two PCR negative controls, were pooled into one multiplex sample and purified using the Minelute PCR purification kit (Qiagen). Two Illumina libraries were built on separate occasions, one containing the samples from the Bahamas, Belize, Jamaica and Turks & Caicos, and one for the samples for the British Virgin Islands, as these samples were collected on a separate expedition. The library of the British Virgin Islands was run along two other libraries (from an unrelated project), equalizing the sequencing depth across all samples by pooling a similar number of samples for each run.

The libraries were built using the NextFlex PCR-free library preparation kit (BIOO Scientific), quantified using the NEBNext qPCR quantification kit (New England Biolabs) and pooled in equal molar concentrations along with 1% PhiX (v3; Illumina) serving as a positive sequencing quality control. The libraries with a final molarity of 10 pM were subsequently sequenced on an Illumina MiSeq platform, in a single MiSeq flow cell using v2 chemistry (2 x 250 bp paired-ends).

4.3.4 **Bioinformatic and statistical analysis**

The bioinformatic analysis was based on the OBITools metabarcoding software suite (Boyer et al. 2016). The pipeline used for data analysis is summarized in Supplementary material 4.2. Paired-end reads were aligned using illuminapairedend and alignments with quality score >30 were kept. The aligned dataset was demultiplexed using ngsfilter, which looks for the known individual sample oligo-tag sequences and additionally remove the metabarcoding primers from every read, leaving only the amplified fragments. A length filter (obigrep) was applied to the assigned reads (300-320 bp) in order to select only the fragments with the correct target size. Reads containing ambiguous bases were also removed. The reads were subsequently dereplicated using obiuniq, grouping all the identical sequences, while keeping track of their abundances. A chimera removal step was performed using the uchime-denovo algorithm (Edgar et al. 2011) implemented in vsearch (Rognes et al. 2016). Sequences were then clustered into Molecular Operational Taxonomic Units (MOTUs), in order to reflect the real species diversity in the samples. These MOTUs were delimited using the step-by-step aggregation clustering algorithm implemented in SWARM 2.0 (Mahé et al. 2015) with a d-value of 13 that has been proven to be most applicable for the Leray fragment (Wangensteen, Palacín, Guardiola, 2017). The SWARM 2.0 algorithm results in variable thresholds for delimiting MOTUs across different branches of the taxonomic tree, which is particularly pertinent with taxonomically highly diverse samples, following the natural organization of the clusters in multidimensional sequence space, and includes one final step that breaks chained MOTUs that could lead to artificial over clustering and hence to an underestimation of MOTU richness values. Taxonomic assignment of the representative sequences for each Molecular Operational Taxonomic Unit (MOTU) was performed using the ecotag algorithm (Boyer et al. 2016) which uses a bespoke reference local database and a phylogenetically-based approach for assigning unmatched sequences to the last common ancestor of the most

closely related sequences in the reference database. We used a bespoke mixed reference COI database (Wangensteen, Palacín, Guardiola 2017), currently containing 191.295 Eukarya sequences, retrieved from the BOLD database (Ratnasingham & Hebert 2007) and the EMBL repository (Kulikova 2004). The database (db_COI_MBPK) is available from http://github.com/metabarpark/Reference-databases. After taxonomic assignment, the final refining of the dataset included taxonomy clustering of MOTUs assigned to the same species and minimal abundance filtering; unassigned MOTUs with less than 2 reads were discarded.

All statistical analyses were performed in R v 3.3.0 (https://www.R-project.org/). Calculations of Bray-Curtis dissimilarity matrices (function vegdist) and group representation in nMDS diagrams (functions ordiellipse and ordispider) were performed using root-root transformed abundance values of frequencies (normalized by dividing MOTU counts by the total counts for each sample). The package MASS (Venables & Ripley 2010) was used for non-metric multidimensional scaling calculations (function isoMDS). SIMPER analyses were performed with the R package Vegan (J. Oksanen *et al.*, 2016) to identify the MOTUs that contribute the most to the differentiation between the sampling locations.

4.4 Results

4.4.1 Read abundances

A total number of 18.745.326 reads was obtained from two Illumina MiSeq runs. After sample assignment, quality and sequence-length filtering and removal of singletons, 2.391.770 reads were left for our 67 samples. The sequencing depth was between 14885-55610 reads per sample. After testing assumptions of normality and homoscedasticity across locations with the Shapiro-wilks (W = 0.98082, n = 68, p-value = 0.38) and Bartlett's tests (K-squared = 1.2203, df = 4, p-value = 0.87) respectively, tests for differences in the number of reads between locations, factorial ANOVA and pair-wise TukeyHSD were performed. We found that there is a difference in the number of reads between locations (F = 3.36, df = 4, p = 0.02), but the pair-wise comparisons showed that this is only significant between Jamaica and the Bahamas (diff = 10390.55, p.adj = 0.03).

After removing bacterial MOTUs, taxonomic assignment resulted in a total of 16.721 MOTUs (Molecular Operational Taxonomic Units), of which 7265 MOTUs could be assigned to phylum level or lower while 9456 MOTUs (56.6%) remained unassigned. The negative controls contained a negligible number of reads. Table 4.1 displays a summary of read statistics per location.

Location	Total reads	Unassigned reads (%)	Metazoa reads (%)	Chordata reads (%)	Fish reads (%)	Total MOTUs	Metazoa MOTUs	Chordata MOTUs	Fish MOTUs
Belize	241238	75.6	12.2	2.05	0.06	2817	458	45	21
Bahamas	430077	65.0	8.64	0.24	0.10	4669	861	45	27
Jamaica	493324	58.6	14.6	0.12	0.005	4971	845	31	9
Turks & Caicos	724252	75.8	14.4	0.84	0.06	7729	1690	52	28
BVI	502879	57.7	17.4	0.10	0.009	6294	1370	61	20

Table 4.1 Summary of the read statistics for all locations

4.4.2 **Patterns of MOTU richness and read abundances**

The number of MOTUs assigned to major eukaryotic groups, at the level of phylum or lower (including 'unassigned Eukarya' and unassigned Metazoa'), for all the samples in each location, are presented (in percentages) in Figure 4.2A. The overall MOTU diversity patterns across all five Caribbean locations are strikingly similar. The 'unassigned Eukarya' group represents by far the largest MOTU richness in every sample, indicating the presence of a significant amount of undescribed eukaryotic biodiversity. Sequences identified as 'other protists' make up the second diverse group. This group is composed of reads, for which it was only possible to determine that these did not originate from organisms belonging to either the Metazoa, Plantae or Fungi Kingdoms, but which subsequently could not be assigned to any higher taxonomic level, adding to the amount of undescribed eukaryotic biodiversity that appears to be present in all five sampling locations. Microscopic organisms such Bacillariophyta (diatoms) and Dinoflagellata also show a relatively high diversity across all samples, as well as Oomyctes and macroscopic seaweeds from the phylum

Rhodophyceae. Species identified as small Metazoa, such as Arthropoda and Annelida, additionally contain a relatively high MOTU richness. And while Chordata are detected in most samples in four out of the five locations (with the exception of Jamaica, where only three samples contain chordate reads), the number of MOTUs assigned to this phylum is very small, indicating low chordate diversity compared to most of the other groups.

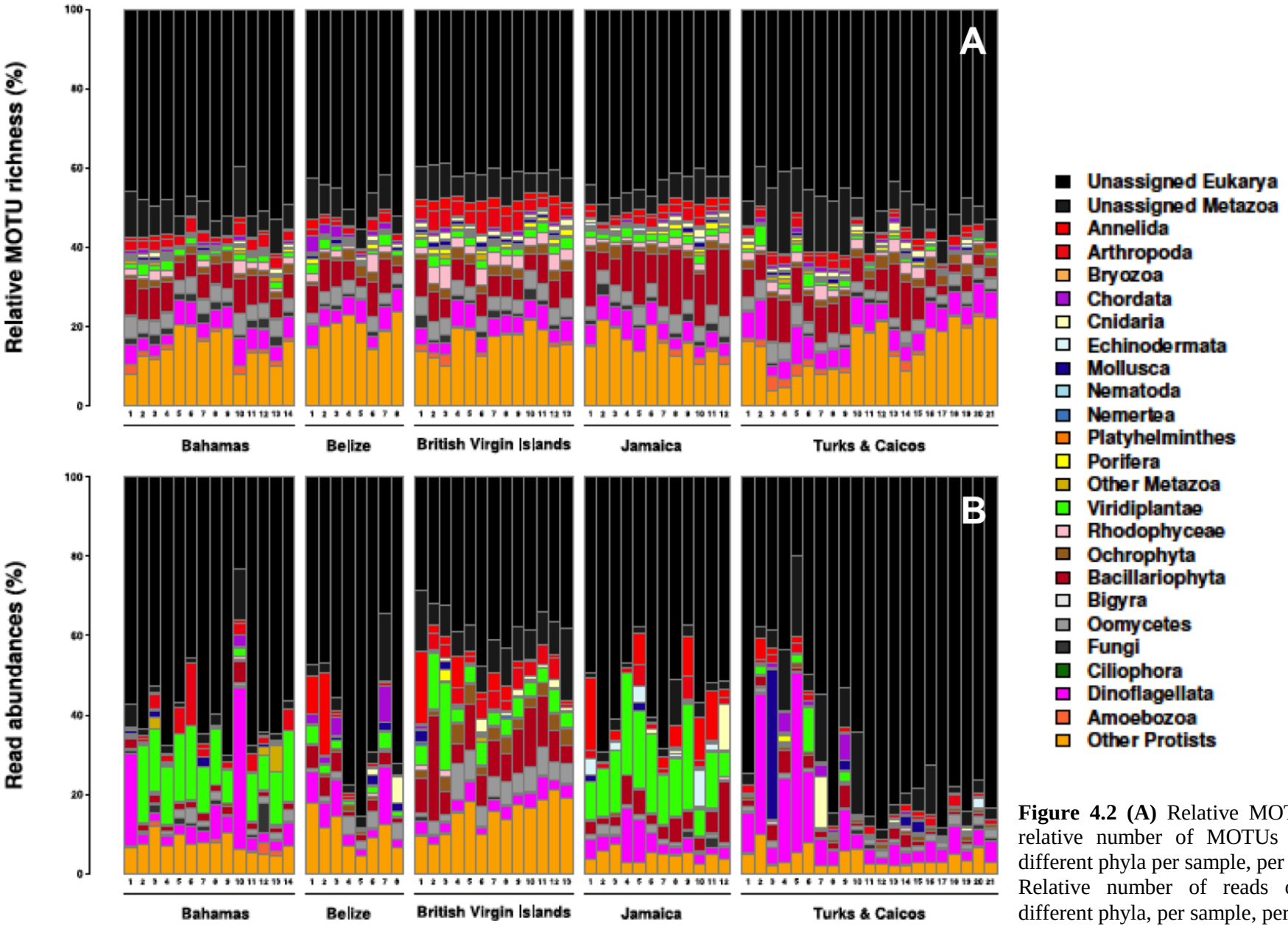


Figure 4.2 (A) Relative MOTU richness; the relative number of MOTUs assigned to the different phyla per sample, per location, and (**B**) Relative number of reads of each of the different phyla, per sample, per location**.**

The abundances of reads assigned to the different phyla in each sample are shown in Figure 4.2 B. Although the patterns of read abundances are much less similar among samples and locations, compared to the patterns of MOTU richness in Figure 3.2 A, it is evident that the 'unassigned Eukarya' not only represent the greatest diversity, but are also dominant in read abundances. The contribution of unknown Eukarya sequences is particularly high in most of the samples from Turks & Caicos. By comparing MOTU richness and read abundances of the different phyla it is apparent that, for example, Bacillariophyta (diatoms) have a relatively high diversity in most of the samples but reads assigned to this phylum are relatively less abundant, except in samples from the British Virgin Islands, where most of the Bacillariophyta reads are assigned to *Rhizosolenia setigera* (34.316 reads), a diatom that is detected only very sparsely in the other four locations (see Supplementary material 4.3 for all the reads per species, per sample). While *Rhizosolenia* species play a significant role in the carbon, silica and nitrogen cycles in the oligotrophic seas, the increase in *Rhizosolenia* is also known to cause fish kills by clogging gills with their hard silica exterior, and from post-bloom anoxia (Pilskaln et al. 2005; Singler & Villareal 2005). MOTU richness and read abundances for Oomycetes, common parasites on marine algae (Li et al. 2010), follow a similar pattern. And while a large part of the Oomycetes reads is assigned to MOTUs that only occur in the British Virgin Islands samples, these could not be taxonomically assigned to any species, genus or family. Additionally, the bulk of unidentified brown algae reads (Phaeophyceae), have been detected in this location. Moreover, both the MOTU for *Bigelowiella natans* (a model organism for the Rhizaria), a protist within the phylum of Cercozoa, and a MOTU belonging to any of the four marine centric diatoms; *Thalassiosira pseudonana, Ditylum brightwellii, Grammonema striatula, Skeletonema menzellii*, were mostly found in the samples from the British Virgin Islands.

Viridiplantae diversity is relatively low overall, compared to for example Bacillariophyta and 'other Protists', however, this group's read abundances are particularly high in most of the samples from Jamaica and the Bahamas, and to a lesser extent in the British Virgin Islands, with (in all three locations) most of the reads being assigned to either *Micromonas commoda* or *M. pusilla*, photosynthetic green algae that are members of the picophytoplankton. Another significant part of the green algae reads are assigned to the MOTU that contains reads from *Mantoniella squamata* and/or *Dolichomastix tenuilepis*, both members of the class Mamiellophyceae. Viridiplantae reads are absent from most of the Turks & Caicos samples, where Dinoflagellata reads are more dominant. Overall, there appears to be a pattern, both location wide, and per sample; where there is a high abundance

of Viridiplantae reads, there is a lower read abundance attributed to Dinoflagellata and vice versa. Hence, while the dinoflagellate MOTU richness is comparable between all samples, the number of reads assigned to dinoflagellates is particularly high in samples with low green algae abundance. Most of these reads are assigned to the class Dinophyceae, with no higher taxonomic classification available. But dinoflagellate reads assigned to *Gymnodinium catenatum*, a planktonic red tide species and paralytic shellfish poison (PSP) producer (Morey-Gaines 1982; Mee et al. 1986), were detected in all but six samples. While for example the peak in dinoflagellate abundance in sample number 10 from the Bahamas is additionally made up of *Gyrodinium impudicum*, which is not associated with PSP (Fraga et al. 1995). Nevertheless, most of the reads in this sample are assigned to either *Heterocapsa circularisquama* or *Heterocapsa triquetra*, both biotoxin producing species (Horiguchi 1995). In many of the samples, reads from the dinoflagellate genus *Symbiodinium* (unicellular algae, colloquially called zooxanthellae), which encompasses the largest and most prevalent group of endosymbiotic dinoflagellates, are also present. The British Virgin Islands is the only location that has a relatively large number of reads assigned to Ochrophyta, a group of mostly photosynthetic heterokonts, with the bulk of the reads originating from brown algae (Phaeophyceae), some of which could be assigned to species level, such as: *Canistrocarpus cervicornis* and *Chattonella subsalsa*, a common bloom species (causing 'brown tides') (Hallegraeff et al. 2003).

Reads from some metazoan phyla, such as the Mollusca, are locally more abundant, in particular in sample 3 from Turks & Caicos. Of those reads, 426 could specifically be assigned to *Chelidonura hirundinina*, an aglajid sea slug, and 88 reads to *Littoraria angulifera*, the mangrove periwinkle, a marine gastropod mollusc belonging to the family Littorinidae. There is also some high local abundance of Cnidaria and Echinodermata in the water column. The peak of Cnidarian reads in sample 7 in Turks & Caicos is solely made up of reads from the Hydrozoan *Cunina fowleri*. Some Scleractinian coral species, such as *Siderastrea radians, Porites astreoides* and *Stephanocoenia michelinii* were also detected. The Echinoderm reads in sample 19 from Turks & Caicos are all assigned to the brittle star *Ophiocoma echinata*. Echinodermata reads in the Jamaican samples are, in addition to *Ophiocoma echinata*, assigned to for example *Lytechinus variegatus* (green sea urchin), *Holothuria impatiens* (bottleneck sea cucumber), and *Actinopyga agassizi* (five-toothed sea cucumber).

While most of the annelid reads could only be assigned to an unidentified Polychaete worm, a large portion of the reads assigned to Annelida in Jamaican sample 1, belongs to

either one of the Polychaetes *Artacama valparaisiensis* or *Polycirrus carolinensis*. Lastly, most of the reads assigned to Chordata are derived from ascidian tunicates (12.458 reads). Some of which could be assigned to species level, such as *Botryllus schlosseri* (golden star tunicate) and the non-native species *Styela plicata* (pleated sea squirt). The large number of chordate reads in Belize samples 3 and 7 and Turks and Caicos 4 and 9, is entirely made up of unidentified species from the tunicate order Enterogona. Most of the remainder of the chordate reads were assigned to Actinopterygii (862 reads) and Chondrichthyes (298 reads), all of which could be taxonomically assigned to species level; 61 teleost and 6 elasmobranch species were detected (table 4.2).

Table 4.2 Teleost and elasmobranch species detected

Species	Common name	Species	Common name
Acanthurus bahianus	ocean surgeonfish	*Sparisoma rubripinne*	redfin parrotfish
Acanthurus chirurgus	doctorfish tang	*Sparisoma viride*	stoplight parrotfish
Acanthurus coeruleus	Atlantic blue tang	*Eucinostomus havana*	bigeye mojarra
Moringua edwardsi	common spaghetti eel	*Haemulon aurolineatum*	tomtate grunt
Atherinomorus stipes	hardhead Silverside	*Haemulon carbonarium*	Caesar grunt
Hypoatherina harringtonensis	reef silverside	*Haemulon flavolineatum*	French grunt
Platybelone argala	keeltail needlefish	*Haemulon plumierii*	white grunt
Strongylura notata	redfin needlefish	*Haemulon sciurus*	blue striped grunt
Tylosurus crocodilus	houndfish	*Lutjanus apodus*	schoolmaster snapper
Parablennius marmoreus	seaweed blenny	*Lutjanus griseus*	mangrove snapper
Acanthemblemaria paula	dwarf spinyhead blenny	*Lutjanus synagris*	lane snapper
Malacoctenus erdmani	blenny	*Ocyurus chrysurus*	yellowtail snapper
Malacoctenus triangulatus	Saddled blenny	*Abudefduf saxatilis*	sergeant major
Carangoides ruber	bar jack	*Chromis cyanea*	blue chromis
Caranx latus	horse-eye jack	*Chromis multilineata*	brown chromis
Brevoortia tyrannus	Atlantic menhaden	*Stegastes adustus*	dusky damselfish
Harengula clupeola	False herring	*Stegastes partitus*	bicolor damselfish
Harengula humeralis	redear herring	*Stegastes planifrons*	threespot damselfish
Anchoa mitchilli	bay anchovy	*Epinephelus guttatus*	red hind
Engraulis eurystole	silver anchovy	*Sphyraena barracuda*	great barracuda
Coryphopterus tortugae	patch-reef goby	*Scomber scombrus*	Atlantic mackerel
Gnatholepis thompsoni	goldspot goby	*Scomberomorus cavalla*	king mackerel
Lythrypnus nesiotes	Island goby	*Scomberomorus regalis*	cero
Ctenogobius saepepallens	dash goby	*Calamus calamus*	saucereye porgy
Apogon aurolineatus	bridle cardinalfish	*Calamus penna*	sheepshead porgy
Clepticus parrae	creole wrasse	*Pseudupeneus maculatus*	Spotted goatfish
Halichoeres bivittatus	slippery dick	*Sphoeroides spengleri*	bandtail puffer
Halichoeres radiatus	puddingwife wrasse	*Carcharhinus perezii*	Caribbean reef shark
Thalassoma bifasciatum	bluehead wrasse	*Negaprion brevirostris*	lemon shark
Xyrichtys martinicensis	rosy razorfish	*Dasyatis americana*	southern stingray
Scarus taeniopterus	princess parrotfish	*Aetobatus narinari*	spotted eagle ray
Sparisoma atomarium	greenblotch parrotfish	*Urobatis jamaicensis*	yellow stingray
Sparisoma aurofrenatum	redband parrotfish	*Ginglymostoma cirratum*	nurse shark
Sparisoma chrysopterum	redtail parrotfish		

4.4.3 **Ordination patterns of community structure**

The ordination of the sampling locations is visualized by a non-metric multidimensional scaling (nMDS) plot in figure 4.3A. Pair-wise comparisons indicate that significant differences in MOTU diversity exist between the five locations (PERMANOVA, Bray-Curtis, $p<0.005$ for all combinations of locations). However, grouped samples from the Bahamas, Jamaica, Turks & Caicos and Belize have partially overlapping inertia ellipses, indicating that at least a proportion of MOTUs is shared between these locations. The samples from Turks & Caicos are slightly more divergent, and the samples from the British Virgin Islands are completely separated from the other four locations. SIMPER (similarity percentage) analysis was used to identify the discriminating taxa between the British Virgin Islands and the other four locations. The MOTU contributing most to this differentiation is assigned to an unidentified Eukaryotic taxon (MOTU number 703; responsible for 17.1% of the differentiation) that is abundant in the four overlapping locations but is very rare in the British Virgin Islands. SIMPER analysis results, showing a list of the 30 most discriminating phyla between the British Virgin Islands and the other four locations, can be found in Supplementary material 4.3. In order to test whether the MOTU diversity in the British Virgin Islands would still be significantly different from the other locations without the abundant Eukaryotic MOTU 703, an nMDS for all the samples without this MOTU is displayed in figure 3B. The British Virgin Islands ellipse remains fully separated from the other locations and pair-wise comparison results remain significant between all locations (Bray-Curtis, $p<0.005$).

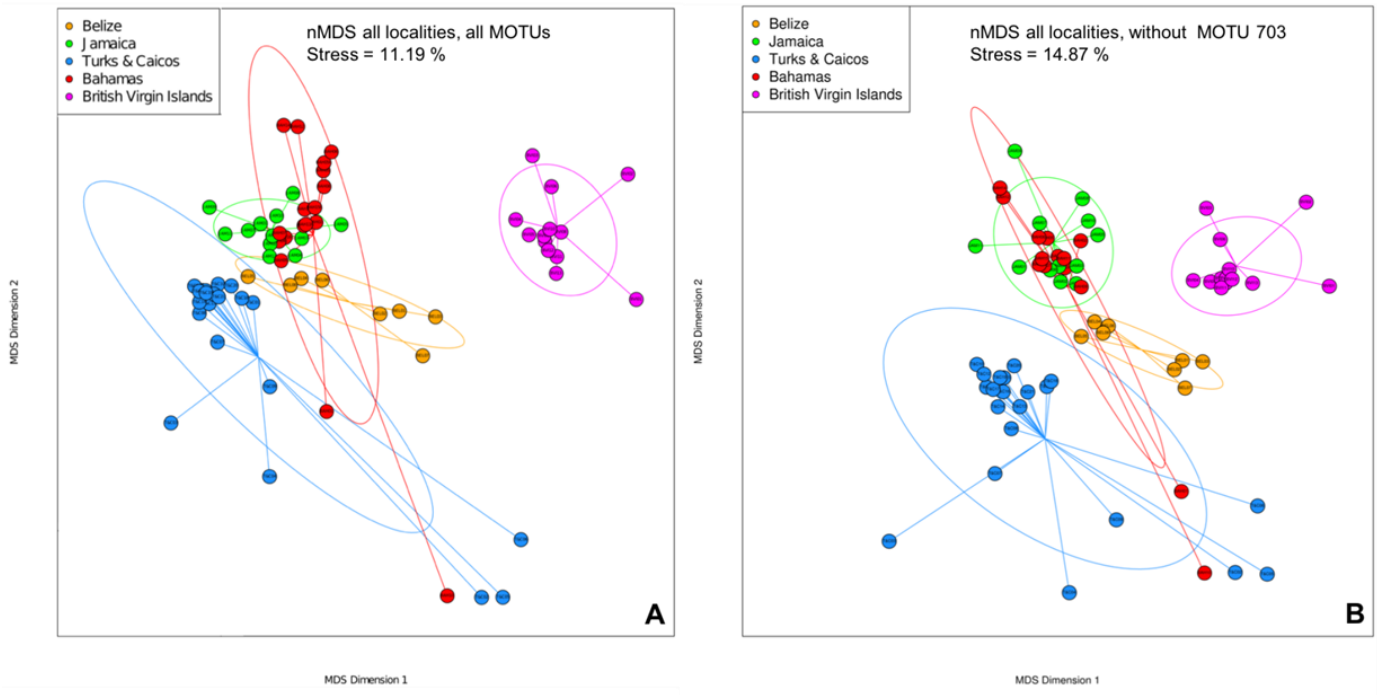


Figure 4.3 Non-metric multidimensional scaling (nMDS) plot (Bray-Curtis, based on relative MOTU read abundances) showing the ordination pattern for the five sampling locations (A) including all MOTUs, and (B) without MOTU 703. Pair-wise comparisons indicate significant differences in MOTU diversity between all the locations, with (A) and without (B) MOTU 703 (Bray-Curtis, $p<0.005$).

For each of the five locations individually, nMDS plots are displayed in figure 4.4, showing the ordination patters within each location. In the Bahamas (Fig. 4.4A), two of the samples are clearly separated from the rest (Bray-Curtis, $p=0.0095$). These are sample number 1 and 10, the two samples with the largest number of dinoflagellate reads (see also Figure 4.2B). Both of these samples were collected from mangrove areas, while the other twelve samples originated from reef sites. SIMPER analysis (Supplementary material 4.4) indicates that an unidentified dinoflagellate MOTU contributes most to the differentiation between the mangrove and reef samples. The second most important MOTU is the same eukaryotic MOTU 703 that is mainly responsible for the separation between the British Virgin Islands and the other locations. As opposed to the dinoflagellate MOTU, this one is much more abundant in the reef samples compared to those from the mangroves. Together with the green algae *Micromonas commoda*, which only appears in the reef samples, these three MOTUs account for almost 29% of the differentiation between reef and mangrove samples.

In Belize, samples were collected from both back (shallow/inside the lagoon) and fore (deeper/towards open water) reef sites, which is reflected in the ordination plot (Fig. 4.4B), in which the samples are separated in fore and back reef ellipses (Bray-Curtis, $p<0.05$). The back-reef samples are characterized by higher numbers of dinoflagellate, Bacillariophyta, Viridiplantae and chordate (tunicate) reads, while the samples from the fore-reef contain more 'unassigned Eukarya', Oomycetes and Cnidaria reads (Fig. 2B). SIMPER analysis shows that the same unknown eukaryote MOTU 703 is responsible for almost 22% of the differentiation between the back and fore reef samples, being more abundant in the fore reef sites. Additionally, the same unknown dinoflagellates are the second most important MOTU, occurring mostly in the back-reef samples. The additional differentiation within the back-reef ellipse can be explained by the fact that these samples were collected from two different back reef sampling sites (Supplementary material 4.1).

The sampling sites in the British Virgin Islands are separated into four different groups (Fig. 4.4C). The samples that were collected directly from the shore (without the use of a boat) are most different from the samples that were collected from the three different reef sites collectively (Bray-Curtis, $p<0.05$). SIMPER analysis indicates that both the unknown eukaryote 703 and dinoflagellate MOTUs contribute most to the dissimilarity in MOTU diversity between the shore and the reef sites. Together with an unidentified mollusc MOTU, they contribute to almost 32% of the total dissimilarity (Supplementary material 4.4).

In Jamaica, samples were collected from two distinct sites, Discovery Bay and Montego Bay, which are grouped separately in Figure 4D. Montego Bay is a relatively large city, home to an airport and a cruise terminal. While Discovery Bay is a small town and receives less tourism. Hence, the separation between the two sites may be related to differences in anthropogenic disturbances and pollution. The unassigned eukaryote MOTU 703 (more abundant in Montego Bay), together with *Micromonas commoda* and an additional unassigned eukaryote MOTU, contribute to 21% of the dissimilarity (Bray-Curtis, $p<0.005$) between the two sites.

From the Turks and Caicos samples, three sampling environments are distinguished by the nMDS (Fig. 4.4D). Similar to the samples from the Bahamas, the samples collected from mangrove areas are grouped separately from the samples collected from reef sites (Bray-Curtis, $p<0.001$), where likewise, the unassigned eukaryote 703 is the most dominant MOTU in the reef samples and is responsible for almost 26% of the dissimilarity between the samples from the two environment types. The unassigned dinoflagellate, being more abundant in the mangrove sites, represents the second most important MOTU. In order to test

whether the separation of samples within the five sampling locations would remain without MOTU 703, nMDS plots for the five locations without 703, were produced and can be found in supplementary material 4.5. Pair-wise comparisons results remained significant for all locations.

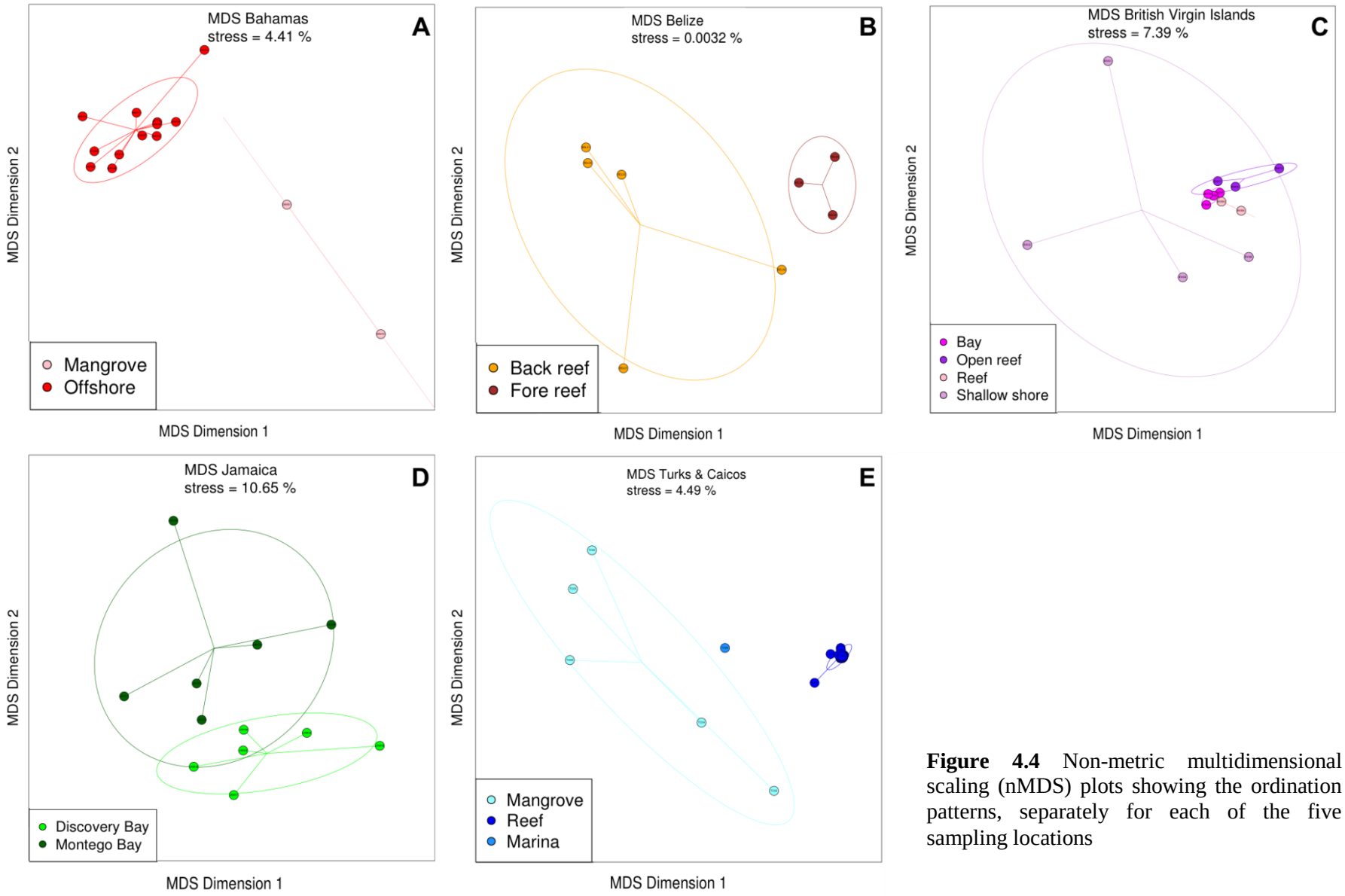


Figure 4.4 Non-metric multidimensional scaling (nMDS) plots showing the ordination patterns, separately for each of the five sampling locations

4.5 Discussion

By applying environmental DNA COI metabarcoding to unfractionated marine water samples, we attempted to characterize marine eukaryotic diversity and spatial patters between and within five Caribbean locations. Our metabarcoding results show a high level of eukaryotic MOTU richness (16.721 MOTUs). However, a disproportionally large amount (56.6%) of all MOTUs could not be assigned to any further taxonomic level beyond Eukaryota, suggesting that the sampled communities host a staggering amount of yet undescribed micro-eukaryotic diversity. Compared to more targeted metabarcoding applications, this approach allowed us to scan biodiversity over a greater taxonomic breadth and additionally, to consider hidden molecular diversity so that ecologically important sequences are included in the overall biodiversity analysis and that the effect of these important taxa on β diversity may still be evaluated, even when these sequences can currently not be taxonomically assigned to a known morphological group or species. This diversity is undetectable with approaches that base taxonomic assignment on database matches with a high percentage of identity, not taking into account the hitherto undescribed component of biodiversity. This is abundantly highlighted by the fact that SIMPER analyses (Supplementary material 4.3) suggest that currently unidentifiable MOTUs may actually be largely responsible for the differentiation of certain communities. Of those analysed in this project, 50% of the differentiations between the different locations/communities is caused by less than 20 MOTUs, of which most have a best id of <0.9 with a species present in our COI database. These dominant MOTUs would have been left out of any taxonomic assignment based on a high level of similarity. Moreover, our data indicate that those taxa that are most important for both differences in read abundances and α-diversity between locations (but also between sites within these locations), are a component of the currently unidentified diversity.

A prime example is the unassigned eukaryote MOTU 703, a taxon that contributes notably to the differentiation between the sampled communities. And while its removal from the analyses does not have a significant effect on the ordination patterns of neither the locations nor the sites within each location, this taxon is responsible for 25% of the variability between the mangrove and reef samples from Turks & Caicos, and it explains 22% of the variability between the fore and back reef sites in Belize. The representative sequence of this MOTU is sequence number 703 (out of 16.721 MOTUs) to appear in the MOTU database, which indicates that it is highly abundant (590.000 reads) and is thus expected to be a relatively important component of the sampled communities. This novel sequence may either belong to

a yet completely undescribed species, or to a species that has already been described morphologically and/or genetically identified with different markers, but for which the COI marker has yet to be described.

Our dataset also contains a relatively high diversity of undescribed animal sequences, indicated by the extent of MOTU richness in the 'Unassigned Metazoa' group (Fig. 4.2 A). It has previously been suggested that indeed many animal lineages remain currently unsampled and/or unsequenced, potentially even harbouring novel groups (López-escardó et al. 2018). Another group in our dataset that is defined by unidentified MOTUs, is the 'Other Protists', a known to be highly diverse, yet heavily undersampled group (Foissner 2008; de Vargas et al. 2015) that may display a wide range of trophic modes (Vaulot et al. 2002) and includes a high diversity of parasites and photosymbiotic taxa (de Vargas et al. 2015).

The British Virgin Islands were sampled two years later than the other locations, which could potentially have played a role in the pronounced separation of this location compared to the other four. However, the samples were collected during the same period (February and March). Moreover, the MOTU explaining 17% of the variability between the British Virgin Islands and the other locations is the unknown eukaryotic MOTU 703, which is a lot less abundant here. Additionally, except for the Bacillariophyta *Rhizosolenia setigera*, most other important MOTUs in the British Virgin Island Samples were also detected in the other four locations.

All sampling methods are subject to methodological limitations, and different sampling methods will capture different subsets of biodiversity (R. Kelly et al. 2017; Shelton et al. 2016). Like more traditional survey methods, eDNA metabarcoding has a certain level of taxonomic selectivity, which may be a result of primer bias. Additionally, when using COI as a metabarcoding marker, the number of species (MOTUs) will often be overestimated due to the existence of 'numts', nuclear sequences of mitochondrial origin (Bensasson et al. 2001; Vamos et al. 2017). Moreover, the use of different metabarcoding markers, with different levels of taxonomic resolution and completeness of reference databases, may produce datasets displaying dissimilar community structure. Consequently, if for this study we had chosen to use a primer set targeting a different marker, such as 18S rRNA (which is often used for studies of planktonic eukaryotes), the dominant unassigned eukaryote MOTU 703 from our study may have been identified to a higher taxonomic level, if this part of its genome has previously been sequenced, while other taxa that were identified down to species level by the Leray-XT primers, may not have been identified at all (López-escardó et al.

2018; de Vargas et al. 2015), due to the limited taxonomic resolution power and the more incomplete eukaryotic database of the 18S marker.

The vast majority of marine life is physically small and dominant in both numbers and diversity (Guil 2011; Snelgrove 1999), consequently in a water sample, eDNA from small eukaryotic organisms far outnumbers that of any vertebrate species, rendering the detection of eDNA from species such as teleosts in unfractionated water samples with a broad-spectrum primer, almost like an eDNA needle in the proverbial haystack. Even more so, at least part of the DNA from microscopic eukaryotes will have originated from entire individuals, as opposed to exclusively extracellular DNA from larger species, potentially drowning out part of the eDNA signal from these larger individuals. Nonetheless, the Leray-XT primer set was capable of detecting not only 61 teleost species but also 6 elasmobranch species.

However, and particularly considering the high levels of diversity present within the sampled habitats, the amplification of predominantly micro eukaryotes compared to larger species, the possibility of taxonomic bias causing inconsistent amplification success rates between different taxa, and the incompleteness of reference databases; the use of a combination of primer sets targeting different metabarcoding markers may be essential for a sufficiently comprehensive community assessment of eukaryotic biodiversity.

While biodiversity loss has been exhaustively documented for macro and mega fauna (which only represent a small fraction of total marine biodiversity), to our knowledge, no research has addressed specifically this issue pertaining to microscopic eukaryotic communities, which is most likely due to the difficulties in characterizing and quantifying the diversity of these communities (Bouchet *et al.*, 2002; Fonseca *et al.*, 2010; Hirai *et al.*, 2015), as many species lack diagnostic morphological features. Despite the limitations of not being able to assign the bulk of detected MOTUs to any taxonomic level beyond Eukaryota, it was still possible to detect differences in community composition between locations and even between sites within those locations. Suggesting that in spite of the current major gaps in reference databases, COI metabarcoding of unfractionated water samples can be useful for describing coarse biodiversity trends. For certain ecological applications, the identification, abundances and patterns of distribution of a particular MOTU may be sufficient, while a scientific designation for that MOTU is yet unavailable (Cordier et al. 2017). However, as species-level analysis of pelagic biodiversity is critical for understanding impacts of climate change, detecting invasive species, and the design of management objectives (Bucklin et al. 2016; Leray & Knowlton 2016), the unravelling of the composition of hidden eukaryotic

diversity and its subsequent description are by no means a minor but essential task, as comprehensive reference databases are critically needed for the taxonomic designation of eukaryotic DNA sequences. And while it has been estimated that between 24 and 98% of marine eukaryotic species are yet to be described (Leray & Knowlton 2016; Goodwin et al. 2017; Mora et al. 2011), the advent of high throughput sequencing and DNA metabarcoding are rendering the huge task of uncovering this hidden diversity using a 'reverse taxonomy' approach (Markmann & Tautz 2005) and in particular taxonomic assignment with COI metabarcoding, less insurmountable. Simultaneously, studies aimed at identifying areas harbouring high numbers of potentially very important microeukaryotes could serve to direct targeted sampling to examine the most abundant microplankton, using powerful microscopy, in order to verify the identity of these taxa which potentially hold great ecological importance.

4.6 Acknowledgements

We thank all staff, students and volunteers from Glover's Reef Research Station, Discovery Bay Marine Laboratory and Field Station, Montego Bay Marine Park, Bimini Biological Field Station (Sharklab), the SFS Center for Marine Studies, Sami Kattan, Diego Camejo, and Ben Williams and Kay van Leuven and the crew of Sunchaser Scuba, for their assistance with sample collection. We also thank Alexandre De Menezes and Ian Goodhead for technical discussions. We are also grateful to Kathryn Matthews and Angelo Villagomez for logistic support. This work was co-funded by the Pew Charitable Trusts, the University of Salford R&E strategy funding and the Unite BVI Foundation.

5 Chapter V

Metabarcoding primers for fish biodiversity assessment: a multi-marker study

Bakker J., Wangensteen O. S., Collins R. A., Soto A. Z., Genner M. J., Mariani S.

Author contributions: S.M., O.S.W., R.A.C., and M.J.G. conceived, designed and coordinated the study; O.S.W., A.Z.S., R.A.C., and M.J.G contributed to fieldwork and sample collection; Laboratory experiments and data analyses were conducted by O.S.W., J.B., R.A.C., and A.Z.S; J.B. wrote the manuscript; all authors read and commented on the manuscript.

5.1 Abstract

Rapid and accurate large-scale species identification and diversity assessment are essential in monitoring and conservation programs for marine species. Environmental DNA (eDNA) metabarcoding is increasingly being applied for the assessment of a diverse range of vertebrate species, including teleosts. However, currently used teleost primers present several shortcomings such as limited taxonomic coverage and limited available reference sequences, which still hinders species-level diversity assessment in taxonomically diverse groups. Here we evaluate the performance of four metabarcoding primers for the assessment of teleost diversity from aqueous environmental DNA samples from UK transitional coastal sites, by specifically focusing on the following variables: (1) completeness of reference database (2) taxonomic coverage (3) taxonomic resolution (4) primer specificity, and (5) overlap between DNA-based identification and morphological surveys. By comparing the data generated by an established 12S primer set with that of three alternative COI primer sets, with different lengths, we test if the enhanced reference databases and taxonomic resolution of COI provide better results. This study shows that eDNA metabarcoding is a promising method for teleost biodiversity assessment and that in future applications, a multi-marker approach will most likely be the most appropriate one. However, significant improvements in both reference databases and taxonomic coverage of primers, are essential.

5.2 Introduction

The development of effective management and conservation strategies for marine fishes depends on accurate population status, biodiversity, and species distribution data. However, detecting species occurrences is often even more challenging in the aquatic environment than on land (Webb & Mindel 2015), and obtaining this data by traditional capture and observation-based sampling methods, is often time-consuming, expensive, and invasive in nature. Moreover, it is subject to intrinsic biases related to catchability and requires taxonomic expertise (Dejean et al. 2012; Taberlet et al. 2012; Takahara et al. 2013; Wheeler 2004).

Currently, environmental DNA (eDNA) metabarcoding of complex samples for aquatic biodiversity assessments is becoming an increasingly popular method for the detection of fish communities in a more resource efficient, comprehensive and non-invasive manner (Bakker et al. 2017; Evans & Lamberti 2017; Yamamoto et al. 2017). The detection of multiple species from eDNA is based on the retrieval of genetic material (e.g. skin cells, metabolic waste, blood), naturally released by organisms in their environment (Ficetola *et al.*, 2008; Taberlet *et al.*, 2012), and the subsequent amplification, sequencing, and taxonomic assignment of this material, through metabarcoding (Ji et al. 2013; Dejean et al. 2012; Thomsen & Willerslev 2015).

It has previously been demonstrated that eDNA metabarcoding has the ability to outperform traditional survey methods for diverse taxa, including fish, both in freshwater (Civade et al. 2016; Deiner et al. 2016; Hänfling et al. 2016; Valentini et al. 2016), and in marine ecosystems (Port et al. 2016; Thomsen et al. 2016; Yamamoto et al. 2017). Not only does high throughput sequencing of eDNA provide a non-invasive and easy to standardize means to rapidly identify multiple taxa without the need for taxonomic identification, it also enables the identification of species that might not be detected using conventional survey methods, such as cryptic species and the juvenile stages of many species, of which the distributions during these phases can only be assumed.

Environmental DNA metabarcoding holds considerable promise to revolutionise our understanding of the spatial and temporal patterns of fish diversity in aquatic environments, particularly in improving estimates of species richness, by revealing the composition of entire fish communities in locations of interest (Lawson Handley 2015; Evans & Lamberti 2017; Yamamoto et al. 2017).

As opposed to the detection of a single species, for biodiversity assessment using eDNA metabarcoding, primers are designed to amplify 'universal' barcoding regions that are evolutionarily conserved across the full spectrum of target species (but still contain enough sequence variability to allow taxonomic resolution of the different species within the targeted community), while having minimal affinity to non-target taxa, such that most of the sequencing depth is dedicated to detecting the species of interest. Hence, the choice of the 'right' primers for the job, is of critical importance.

Several markers, of different lengths, have been proposed for both the detection of individual fish species, and for the characterization of fish communities from aqueous eDNA (Ficetola et al. 2010; Kelly, J. a. Port, et al. 2014; Leray et al. 2013; Miya et al. 2015; Riaz et al. 2011; Stoeckle et al. 2017; Thomsen, Kielgast, Lars L. Iversen, et al. 2012; Valentini et al. 2016), however, consensus on an optimal generic fish marker has yet to be reached (Shaw et al. 2016), and studies comparing different markers on the same samples, particularly pertaining to marine species, are scarce.

Currently, the mitochondrial cytochrome oxidase I (COI) and the 12S marker gene regions are the most widely used in fish DNA barcoding studies (Hardy et al. 2011; Shaw et al. 2016). The (COI) barcode region (Hebert, Ratnasingham, et al. 2003), is one of the most commonly sequenced regions for analysis of species diversity among marine animals (Bucklin et al. 2011; Bucklin et al. 2016), due to its high taxonomic resolution resulting from high mutation rate. COI variability is large enough to allow the discrimination of closely related species in most groups, and can even inform on intraspecific variation associated with geographic structure (Bucklin et al. 2011). Additionally, the availability of the large Barcode of Life Data (BOLD) system greatly facilitates taxonomic assignment (Clarke et al. 2017; Hebert, Cywinska, et al. 2003; Ratnasingham & Hebert 2007). However, since the COI region shows high codon degeneracy ('third codon wobble') throughout its sequence, it lacks highly conserved primer-binding sites (potentially causing taxonomic bias through primer-template mismatches when targeting genetically diverse taxonomic groups), making the design of universal primers, such as is essential for biodiversity metabarcoding studies, very difficult (Clarke et al. 2014; Deagle et al. 2014; Pinol et al. 2015; Sharma & Kobayashi 2014). Instead, it has been argued that the use of mitochondrial non-coding, ribosomal markers, with more conserved primer binding regions, such as the 12S region, may be more appropriate (Clarke et al. 2014; Deagle et al. 2014; Miya et al. 2015; Yang et al. 2014). Nonetheless, due to the higher level of sequence conservation, ribosomal markers often have

limited taxonomic resolution (potentially underestimating species diversity within a community), and additionally, have less exhaustive reference databases, compared to COI.

Regardless of marker choice, whether a target species in a complex mixed eDNA sample can be detected, will depend on a number of factors (Fig. 5.1) (1) completeness of the reference database; if a reference sequence for this species is available, (2) taxonomic coverage of the chosen primer set; if the species-specific barcode sequence can be amplified, (3) taxonomic resolution of the barcoding marker; if the marker sequence is able to characterize the species (and not just the genus or family it belongs to), and (4) the taxonomic specificity of the primer set; if the produced reads of all target species are of sufficient abundance, compared to non-target taxa that may be present in the same sample.

In this study, we evaluate and compare the theoretical and practical performances of four primer sets, targeting both the mitochondrial COI and 12S regions (Table 5.1), for the diversity assessment of teleost fish communities from coastal and transitional waters in the United Kingdom. We used *in silico* PCR (Ficetola et al. 2010) to compare taxonomic coverage and resolution of novel COI fragments of two lengths (SeaDNA-Short 55 bp and SeaDNA-Mid 130 bp), specifically designed for the amplification of teleosts. Another, more universal, COI primer set (Leray-XT, 313 bp) targeting eukaryotic diversity (Wangensteen, Palacín, Guardiola 2017), and one established 12S, teleost specific, primer set (MiFish, 170 bp) (Miya et al. 2015). Subsequently we compared their efficiency in teleost eDNA amplification, *in vitro*, in order to determine primer efficiency in the amplification and characterization of teleost diversity in natural eDNA samples, from four locations in North Sea and English Channel coastal sites in Britain (Fig. 5.2). Additionally, an important aspect in testing the potential of eDNA methods to be applied as a tool in monitoring marine species, is to survey the natural environment using both traditional and eDNA methods in concert. Thus, the various primers are also compared in terms of fish detection compared to concurrent morphological surveys.

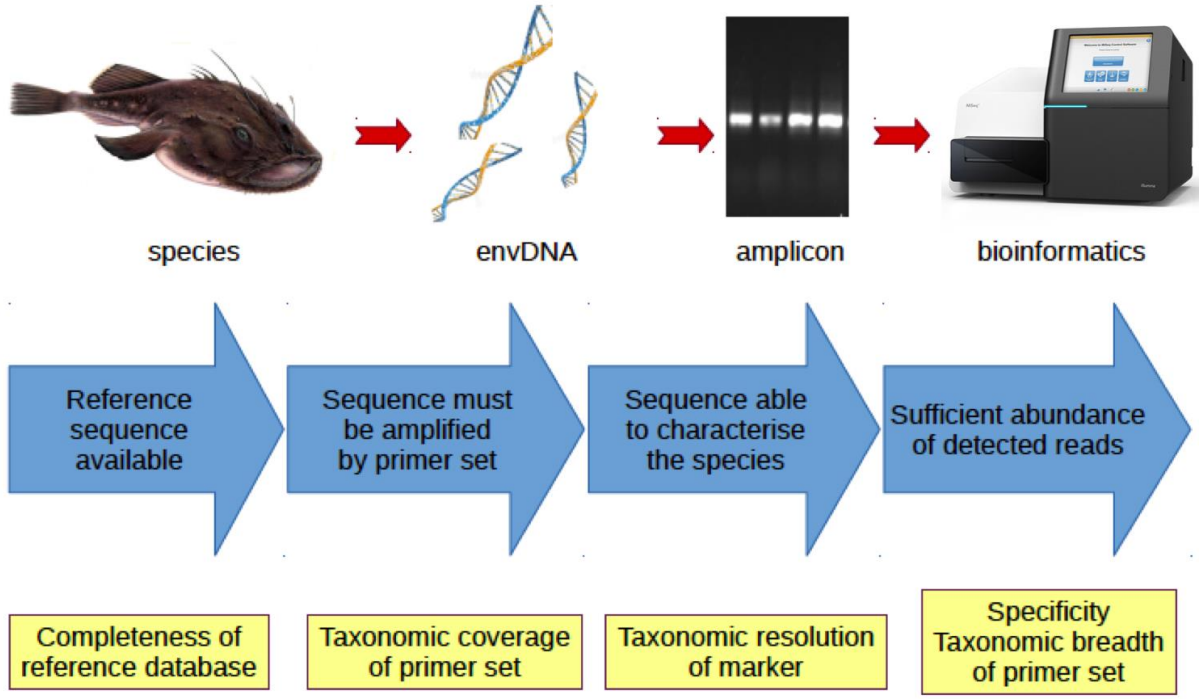


Figure 5.1 Factors determining whether the DNA of a species, present in a complex, mixed eDNA sample, will be detected by metabarcoding

5.3 Material and methods

5.3.1 Water sampling

A total of 15 sites were sampled, in four coastal British locations. These included: the Tees Estuary, two sites within the Esk Estuary, the Test Estuary and Whitsand Bay, sampled between October and November of 2016. The former three are estuarine sites, while the latter is a coastal shelf area. These locations are major habitats with importance for fisheries, environmental monitoring and conservation. A map of sampling locations is shown in Fig. 5.2. Three replicate samples per site, consisting of 2 litres of water, were collected by hand (while wearing disposable gloves) in sterile collection bottles. In order to minimize the amount of plankton (non-target taxa) in the samples and to remove algae and sediment (reducing the severity of filter clogging) prior to passing the samples through the filters, collection bottles were covered with a 250 µm pore size nylon mesh pre-filter, attached to the neck of the bottle by an elastic band. The mesh was discarded after the water sample was collected.

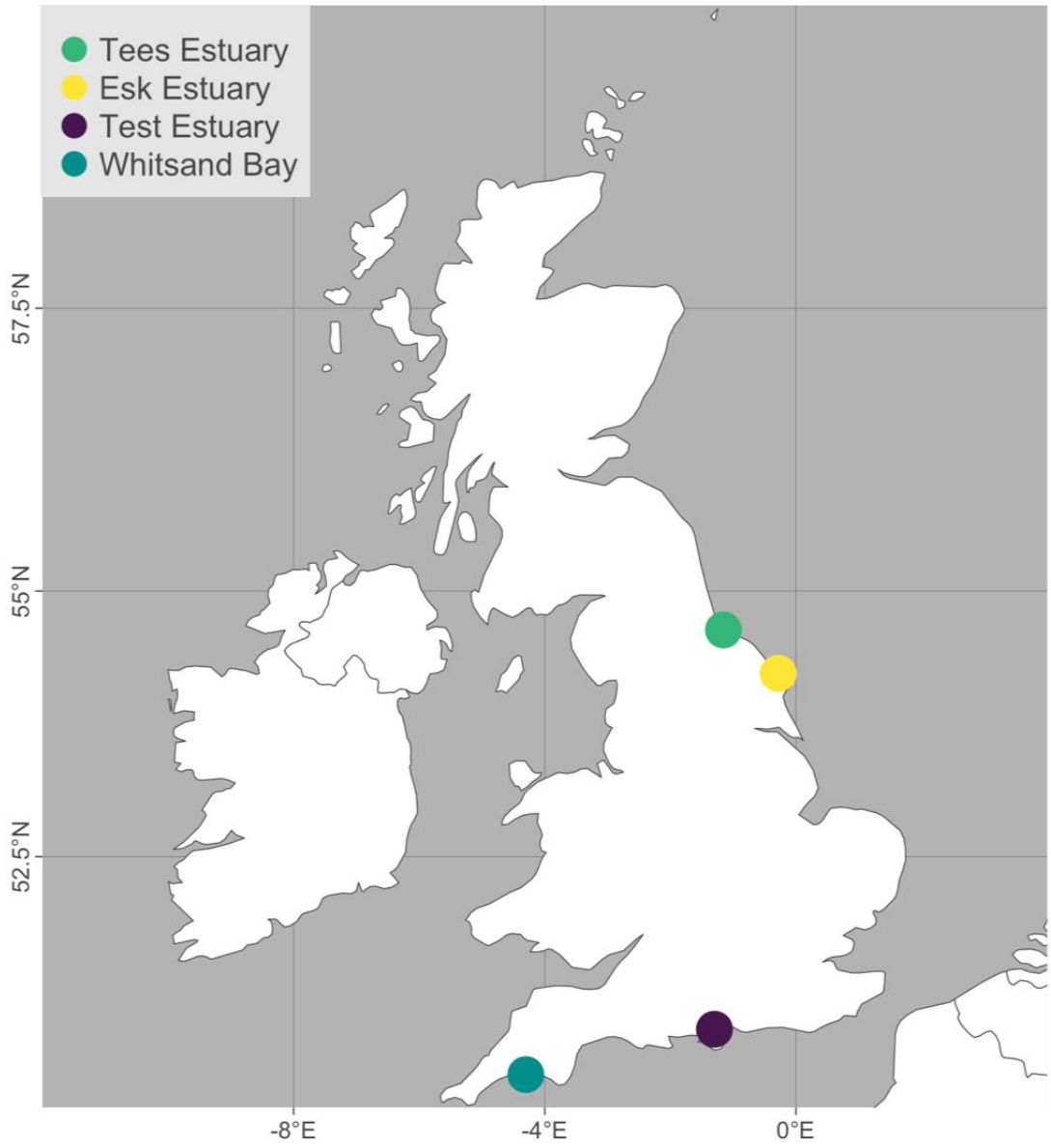


Figure 5.2 Map of the United Kingdom, showing the four sampling locations. The Tees Estuary, the Esk Estuary, the Test Estuary and Whitsand Bay.

5.3.2 **Fishing**

Fish sampling in the Esk estuary was done by two replicate fyke nets (Esk-fyke) and two replicate beach-seine nets (Esk-seine), during October 2016. At the Tees sampling site, fish surveys were conducted by both two beach-seine nets and two shallow beam trawls, also during October of 2016. During November of the same year, impingement monitoring of fish species was conducted in the Test Estuary and four otter trawls as described in (Sims et al. 2004), were conducted in Whitsand Bay. The variety of fishing techniques used in the

different sampling locations are part of the currently ongoing monitoring programmes implemented by our local collaborating organizations and stakeholders; the Environment Agency, PISCES Conservation Ltd. And the Marine Biological Association.

5.3.3 **Sample processing and DNA extraction**

After collection, the water samples were individually covered, put into individual sterile plastic bags, and stored in a cooler box filled with ice, while being transported back to the dedicated controlled eDNA laboratory facilities at the University of Salford. Within five hours after collection, each 2-litre sample was filtered through a sterile 0.22 µm Sterivex™-GP filter (polyethersulfone (PES), Merck Millipore, Germany) using a 100 mL polypropylene syringe (Merck Millipore, Germany). After filtration, the filters were stored at -20 ºC, prior to extraction. DNA was extracted from the filters with the DNeasy PowerSoil DNA Isolation Kit (Qiagen), following the manufacturers' protocol, with the addition of an initial 2-hour agitation step, during which the membrane filters were placed in tubes with lysis buffer (C1) and garnet beads from the PowerWater Isolation kit. The tubes were subsequently placed in an orbital shaker at low speed, at 65 ºC, in order to promote the release of DNA from the membrane. Filtration blank controls were processed in parallel. Purified extracts were assessed for DNA concentration in a Qubit fluorometer (Thermo Fisher Scientific). Disposable nitrile gloves were worn during all stages of sample and filter processing and laboratory equipment and surfaces were cleaned using a 50% bleach solution.

5.3.4 **Primer Design**

We designed two new metabarcoding primers, SeaDNA-short and SeaDNA-mid, specifically targeting teleosts. These primers were designed using a dataset of complete COI genes for fishes. A total of 2,317 complete fish mitochondrial genomes were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genome/browse/) in November 2016. The COI genes were extracted using a hidden Markov model in the program HMMER v3.1 http://hmmer.org/ (Eddy 1998), taxonomic annotations were added using rfishbase v2.1.2 (Boettiger et al. 2012), and the sequences were aligned using MAFFT v7.123 (Katoh & Standley 2013). Primers were then designed manually in Geneious 8.8.1 (Kearse et al. 2012), with the assistance of Primer3 (Untergasser et al. 2012) and the sliding window functions in spider

v1.3.0 (Boyer *et al.*, 2012; Brown *et al.*, 2012). The new primer sets are both internal to the Folmer fragment (Folmer et al. 1994), which is commonly used for DNA barcoding.

5.3.5 **Library preparation and sequencing**

Four different metabarcoding markers were amplified, for which the details are presented in table 5.1. To all four primers, 8-base sample specific oligo-tags, differing in at least 3 bases were attached (Guardiola et al. 2015). In order to increase variability of the amplicon sequences, a variable number (2, 3 or 4) of fully degenerate positions (Ns) was added at the beginning of each primer (Wangensteen & Turon 2017). The PCR conditions for the 12S MiFish primers followed (Miya et al. 2015), and for the Leray-XT primers, (Wangensteen, Palacín, Guardiola 2017). For PCR amplification with the newly designed SeaDNA-Short and SeaDNA-Mid primers, a two-step protocol was used, attaching the 8-base tagged primers, after an initial amplification, in the second PCR. The mix recipe for the first PCR included AmpliTaq Gold DNA polymerase, with 1 µl of each 5 µM forward and reverse primers, 0.16 µl of bovine serum albumin and 10 ng of purified DNA in a total volume of 20 µl per sample. The recipe for the second PCR was identical, except for the primers now being the forward and reverse 8-base tagged primers. For the first stage PCR, the profile included an initial denaturing step of 95 ºC for 10 minutes, 40 cycles of 94 ºC 30 sec, 47 ºC 45 sec and 72 ºC 30 sec and a final extension step of 72 ºC for 5 minutes. The profile for the second stage PCR was identical, except for the annealing temperature being 50 ºC instead of 47 ºC.

All PCR amplifications were done in duplicate reactions to minimize PCR bias. The quality of all amplifications was assessed by electrophoresis, running the products through a 1.5% agarose gel stained with Gel Red (Cambridge Bioscience) and visualized on a UV light platform. Between the first and second PCR step, amplicons were purified using MinElute PCR purification columns (www.qiagen.com) and diluted ten times prior to being used as a template for the second PCR. After the second PCR, all tagged amplicons were pooled by marker, purified using MinElute columns and each pool was eluted in a total volume of 45 µl, in order to concentrate the amplicons approximately 15 times, for NGS library preparation.

Libraries (one for each marker) were built using the ligation-based NetFlex PCR-free library preparation kit (BIOO Scientific). The libraries were quantified using the NEBNext qPCR quantification kit (New England Biolabs) and pooled in equimolar concentrations along with 1% PhiX (v3, Illumina) serving as a positive sequencing quality control. For each primer, the

15 samples were run using two PCR duplicates, along with one filtration and one PCR blank. The libraries were sequenced on an Illumina MiSeq platform, using V3 chemistry (2x75bp paired-end run) for the SeaDNA-Short library, which was run along two other libraries (from an unrelated project). For the MiFish and SeaDNA-Mid libraries V2 chemistry (2x150bp paired-end run) was used, and these were sequenced in the same run. The Leray-XT library was run using V2 (2x250bp paired-end run) chemistry along with one different library (from an unrelated project). Sequencing depth for all libraries was approximately similar.

A dedicated controlled eDNA lab, with separate rooms designated for the physical separation of eDNA extraction, pre-PCR preparations and post-PCR procedures, was used for all laboratory work. Moreover, to identify potential contamination, every library included one filtration blank (DNA extraction from a Sterivex filter after passing 2L of commercial drinking water), and one PCR blank.

Table 5.1 Details of the PCR primer sets used in this study. Amplicon lengths are given both excluding and including primer sequences, individual sample tags, and leading N's.

Name	Locus	Primer sequence (5'-3')	Amplicon length (bp)	Full length; amplicon + primers + sample tags + leading N's (bp)	Reference
SeaDNA-Short	COI	GGAGGCTTTGGMAAYTGRYT GGGGGAAGAARYCARAARCT	55	236	This study
SeaDNA-Mid	COI	GGAGGCTTTGGMAAYTGRYT TAGAGGRGGGTARACWGTYCA	130	312	This study
Leray-XT	COI	GGWACWRGWTGRACWITITAYCCYCC TAIACYTCIGGRTGICCRAARAAYCA	313	510	Wangensteen *et al*. (2017)
MiFish	12S	GTCGGTAAAACTCGTGCCAGC CATAGTGGGGTATCTAATCCCAGTTTG	165	345	Miya *et al*. (2015)

5.3.6 **In silico evaluation of the primer sets**

In order to compare taxonomic coverage and species-level resolution of all four primer sets, they were evaluated *in silico* against 160 teleost species that are found in UK transitional

coastal waters. A species list, provided by the Water Framework Directive, United Kingdom Technical Advisory Group, is available in Supplementary Material 5.1. Available full COI and 12S sequences for these species were downloaded from Genbank and the ability of each primer set to amplify the different species (taxonomic coverage) was assessed using ecoPCR and the ecotaxstat function (Ficetola et al. 2010), allowing 3 mismatches per primer, and no mismatches in the two base pairs at the 3' end. The program ecoPCR uses a pattern-matching algorithm to identify sequences within a database that can be amplified with a given primer pair by constraining the relative orientation of and maximum distance between primer-binding sites, as well as the number of mismatches between primer and target sequences (Ficetola et al. 2010). This approach could not be used directly for partial COI sequences (corresponding to the standard Folmer barcoding region) to evaluate the Leray-XT primer set, since most barcode sequences in the Genbank and BOLD databases usually lack the reverse primer binding sequence (as it lays just outside the Folmer fragment) (Wangensteen et al. 2018). Thus, for those teleosts for which no complete COI sequences were available, ecopcr and ecotaxstat were ran against sets of COI barcode sequences with an artificial jgHCO2198-matching sequence attached to the 3' end. This made it possible to test the coverages of the internal forward primers and to compare the performance of the Leray-XT primer set with the other three primer sets, although the taxonomic coverage produced by this primer set could be overestimated using this approach, as it does not account for potential mismatches in the reverse primer region of the partial COI sequences. To check the taxonomic resolution of the markers, the ecotaxspecificity function was used (Ficetola et al. 2010). This function calculates the ratio of the amplified species that can be unambiguously identified by the amplified fragment.

5.3.7 **Bioinformatic and statistical analysis**

The metabarcoding pipelines were based on the OBITools software suite (Boyer et al. 2016). Quality of the raw reads was assessed using FastQC and the length of raw reads was trimmed to a median Phred quality score >30, after which paired-end reads were assembled using illuminapairedend. The reads with alignment quality scores >40 were kept and the resulting dataset demultiplexed using ngsfilter. A length filter (obigrep) was applied to the aligned reads (45-65 bp for SeaDNA-Short, 120-140 bp for SeaDNA-Mid, 303-323 bp for Leray-XT and 155-185 for MiFish) in order the select only the fragments with the correct target size. Reads containing ambiguous bases were also removed. The reads were then dereplicated

using obiuniq, and chimeric sequences were detected and removed using the uchime-denovo algorithm (Edgar et al. 2011) implemented in vsearch (Rognes et al. 2016). The MOTUs were subsequently delimited using the step by step aggregation clustering algorithm implemented in SWARM 2.0 (Mahé et al. 2015) with a d-value of 13 for Leray-XT (Wangensteen et al. 2018), d=3 for Miya, d=2 for SeaDNA-short, d=6 for SeaDNA-Mid. The values for this parameter were chosen taking into consideration the natural variability and the length of the different fragments, and the homogeneity of taxonomic assignment of the clustered sequences was checked for selected MOTUs, in order to validate the used d values. The SWARM 2.0 algorithm results in variable thresholds for delimiting MOTUs across different branches of the taxonomic tree, following the natural organization of the clusters in multidimensional sequence space, and includes one final step that breaks chained MOTUs that could lead to artificial over clustering and hence to an underestimation of MOTU richness values. Taxonomic assignment of the representative sequences for each MOTU was performed using the ecotag algorithm (Boyer et al. 2016), which uses a bespoke reference local database and a phylogenetically-based approach for assigning unmatched sequences to the last common ancestor of the most closely related sequences in the reference database. A COI database (Wangensteen, Palacín, Guardiola, 2017) containing 191.295 Eukarya sequences, retrieved from the BOLD database (Ratnasingham & Hebert 2007) and the EMBL repository (Kulikova 2004), was used for the Leray-XT primers. Bespoke teleost databases for the MiFish 12S (6868 sequences), SeaDNA-Mid COI (67070 sequences), and SeaDNA-Short COI (259696 sequences) primer sets were built with sequences retrieved from Genbank. Reference databases are publicly available from: http://github.com/metabarpark/Reference-databases. After taxonomic assignment, the final refining of the datasets included taxonomic clustering of MOTUs assigned to the same species and minimal abundance filtering; unassigned MOTUs with less than 5 reads for SeaDNA-Mid, less than 3 reads for SeaDNA-Short, 2 reads for Leray-XT and 60 reads for MiFish, were discarded. The minimum read abundance cut-off rates per marker were optimized to reduce the number of false positive MOTUs, considering the number of teleost reads generated by each of the markers. After taxonomic assignment, the resulting datasets from the PCR duplicates from each sample, were combined in order to reduce the effect of PCR amplification bias and sequencing errors (Alberdi et al. 2017; Burgar et al. 2014; Leray & Knowlton 2015) in order to maximize diversity detection. The pipelines used for data analysis, for both metabarcoding markers, are summarized in Supplementary material 5.2. All statistical analyses were performed in R v 3.3.0 (https://www.R-project.org/). Package vegan

(function speccacum) (Oksanen et al., 2016) was used for obtaining MOTU accumulation curves. Custom R scripts are publicly available from http://github.com/metabarpark.

5.4 Results

5.4.1 *In silico* evaluation of the primer sets

We have tested the theoretical performance of the COI and 12S primer sets, by computationally evaluating them against a list of 160 teleost species found in estuarine and transitional waters in the UK. Based on the four factors determining species detectability as shown in Figure. 5.1, the theoretical performances of each of the four primer sets are displayed in Table 5.2. All three COI primer sets, have a reference sequence available for 94.4% of the 160 species, while for the 12S primer set, a reference sequence is available for only 51.9% of the 160 different teleost species. Additionally, the COI primer sets have relatively high taxonomic coverage and taxonomic resolution, compared to the 12S primers set. However, the performance in taxonomic specificity of the 12S primer set, is superior compared to all three COI primer sets, indicating that 77% of the amplified reads will belong to teleost species, as opposed to non-target taxa. Thus, based on *in silico* analysis, the 12S marker has an incomplete reference database, moderate taxonomic coverage, moderate taxonomic resolution, but a very high specificity for the amplification of teleost DNA. Hence, it is expected to produce a high number of teleost reads, but it will possibly yield a relatively low number of detected species, with moderately accurate species identification compared to the COI markers, which have a much more exhaustive reference database, good taxonomic coverage, very good taxonomic resolution, but very low taxonomic specificity. Accordingly, the COI primer sets are expected to produce a very low number of teleost reads (compared to reads from non-target taxa), but from a more diverse assemblage of species. Moreover, they are expected to show highly accurate species identification ability compared to the 12S primer set.

Table 5.2 Results from the in silico analysis, comparing the theoretical performance of the four primer sets in the detection of 160 teleost species from UK transitional waters.

Primers	**Reference database completeness**	**Taxonomic coverage**	**Taxonomic resolution**	**Specificity**
12 S Miya MiFish 170 bp	83 spp.: **51.9%**	69 spp.: **83.1%**	62 spp.: **89.8%**	High: **77%** teleost reads
COI Leray-XT 313 bp	151 spp.: **94.4%**	144 spp.: **95.5%**	142 spp.: **98.6%**	Very low: **0.05%** teleost reads
COI SeaDNA-Mid 130 bp	151 spp.: **94.4%**	128 spp.: **84.8%**	108 spp.: **84.4%**	Low: **1.2%** teleost reads
COI SeaDNA-Short 55 bp	151 spp.: **94.4%**	147 spp.: **97.3%**	121 spp.: **82.3 %**	Low: **1.2%** teleost reads

5.4.2 **Teleost detection**

A total of 15 samples, from 4 locations (5 sites), in 4 different amplicon libraries, were sequenced using three separate Illumina MiSeq runs. The read statistics for all four primer sets after sample assignment, quality and sequence-length filtering, and combining the results from the PCR duplicates from each sample, are displayed in Table 5.3. No teleost reads were detected in the negative controls. Overall, 60 teleost MOTUs were detected by eDNA metabarcoding, of which 36 species were exclusively detected by eDNA. In total, 41 different species were detected by the morphological surveys, of which 18 species were not detected by eDNA. A total of 27 species was detected by both methods (Figure 5.3).

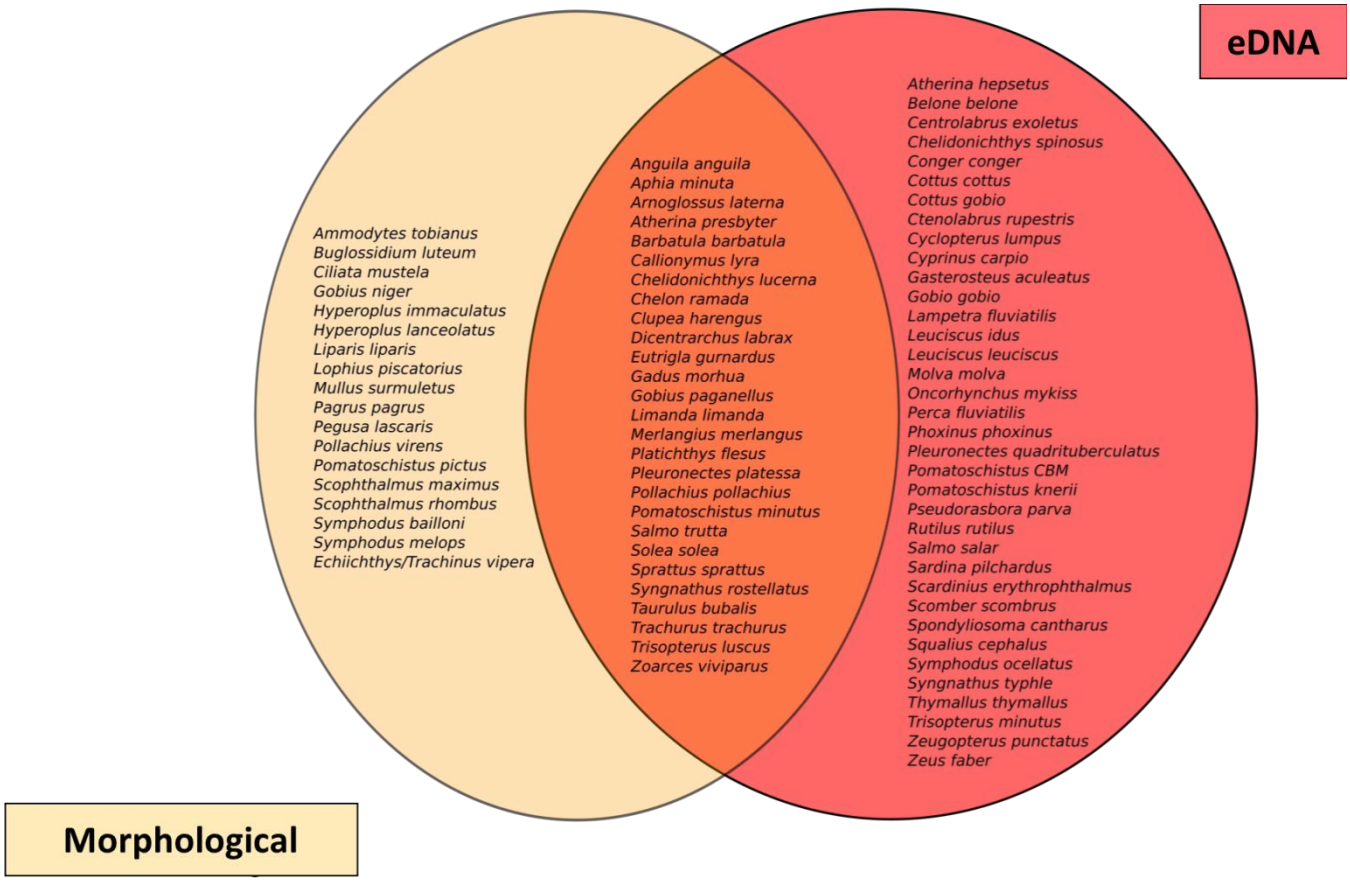


Figure 5.3 Comparison of the different species detected by morphological surveys and by eDNA metabarcoding. The overlap represents the 27 species detected by both methods.

Table 5.3 Sequencing read statistics for the four primer sets

Primer set	Total reads	Teleost reads	Teleost reads (%) of total reads	Teleost MOTUs	Metazoa reads	Metazoa reads (%) of total reads	Other Eukarya reads	Unidentified Eukarya reads	Other Eukarya reads (%) of total reads	Bacteria reads	Bacteria reads (%) of total reads
COI Leray-XT	3062194	1554	0.051	32	15968	0.521	1242435	407511	40.573	1396280	45.597
COI-SeaDNA-Mid	8201135	97589	1.19	17	276619	3.373	5313981	1308357	64.796	1302178	15.878
COI-SeaDNA-Short	4447441	54392	1.223	18	103442	2.326	1275366	553237	28.676	2515396	56.558
12S Miya MiFish	2312505	1795728	77.653	33	2295310	99.256	17195	0	0.744	0	0

As shown in Figure 5.4, there is only a relatively small overlap in species detection between the four different primer sets. Eight species are detected by all four primers, six marine species, *Aphia minuta* (transparent goby), *Gasterosteus aculeatus* (three-spined stickleback), *Oncorhynchus mykiss* (rainbow trout), *Salmo salar* (Atlantic salmon), *Salmo trutta* (brown trout) and *Trisopterus luscus* (whiting-pout), and two fresh water species *Barbatula barbatula* (stone loach) and *Gobio gobio* (gudgeon). While a total number of 60 species was detected, half (30 species) was detected by only one of the four primer sets. Leray-XT uniquely amplified 10 species, while 13 species were only amplified by MiFish, 4 by SeaDNA-Mid and 3 by SeaDNA-Short. If one would have to choose a combination of two primer sets, Leray-XT and MiFish combined, detect a total of 51 different teleost MOTUs. However, SeaDNA-Short is the only primer set that has detected *Pseudorasbora parva* (indicated by a blue star in Fig. 5.4), the topmouth gudgeon (or stone moroko), which is native to Asia, but has been introduced and is now considered an invasive species in Europe (Britton et al. 2010; Pindera et al. 2005). Moreover, 5 out of the 33 MOTU's that were detected by the MiFish primer set, were taxonomically assigned, with 100% accuracy, to species that do not occur in UK waters. Both *Chelidonichthys spinosus* (spiny red gurnard) and *Pleuronectes quadrituberculatus* (Alaska plaice) are Pacific species (red stars in Fig. 5.4) while their Atlantic counterparts *Chelidonichthys cuculus* (red gurnard) and *Pleuronectes platessa* (European plaice), indeed do occur in the North Sea. Moreover, three MOTU's delimited by the MiFish primer set, have been taxonomically assigned to Mediterranean species, *Atherina hepsetus* (Mediterranean sand smelt), *Pomatoschistus knerii* (Kner's goby) and *Symphodus ocellatus* (ocellated wrasse) (green stars in Fig. 5.4). Most likely, these sequences are derived instead from similar species occurring in the North Sea; *Atherina presbyter* (sand smelt), *Pomatoschistus minutus* (sand goby) and *Symphodus melops* (corkwing wrasse). The incorrect taxonomic assignment to sister species is probably caused by the fact that many sequences deposited in the 12S teleost database are derived from Pacific counterparts while the sequences of the European species are absent from the reference database. Details for the number of reads per species, per location, for each primer set, are available in Supplementary material 5.3.

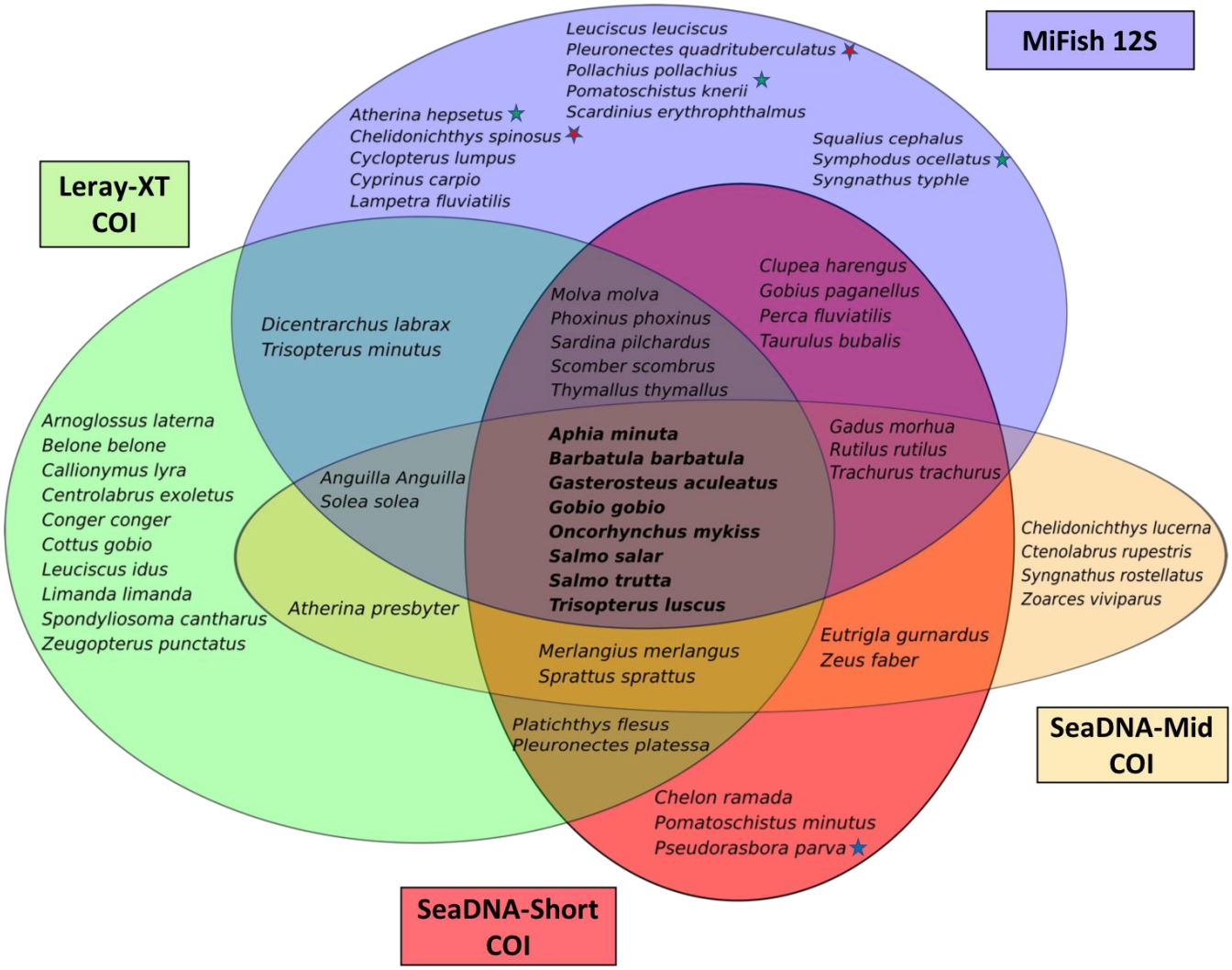
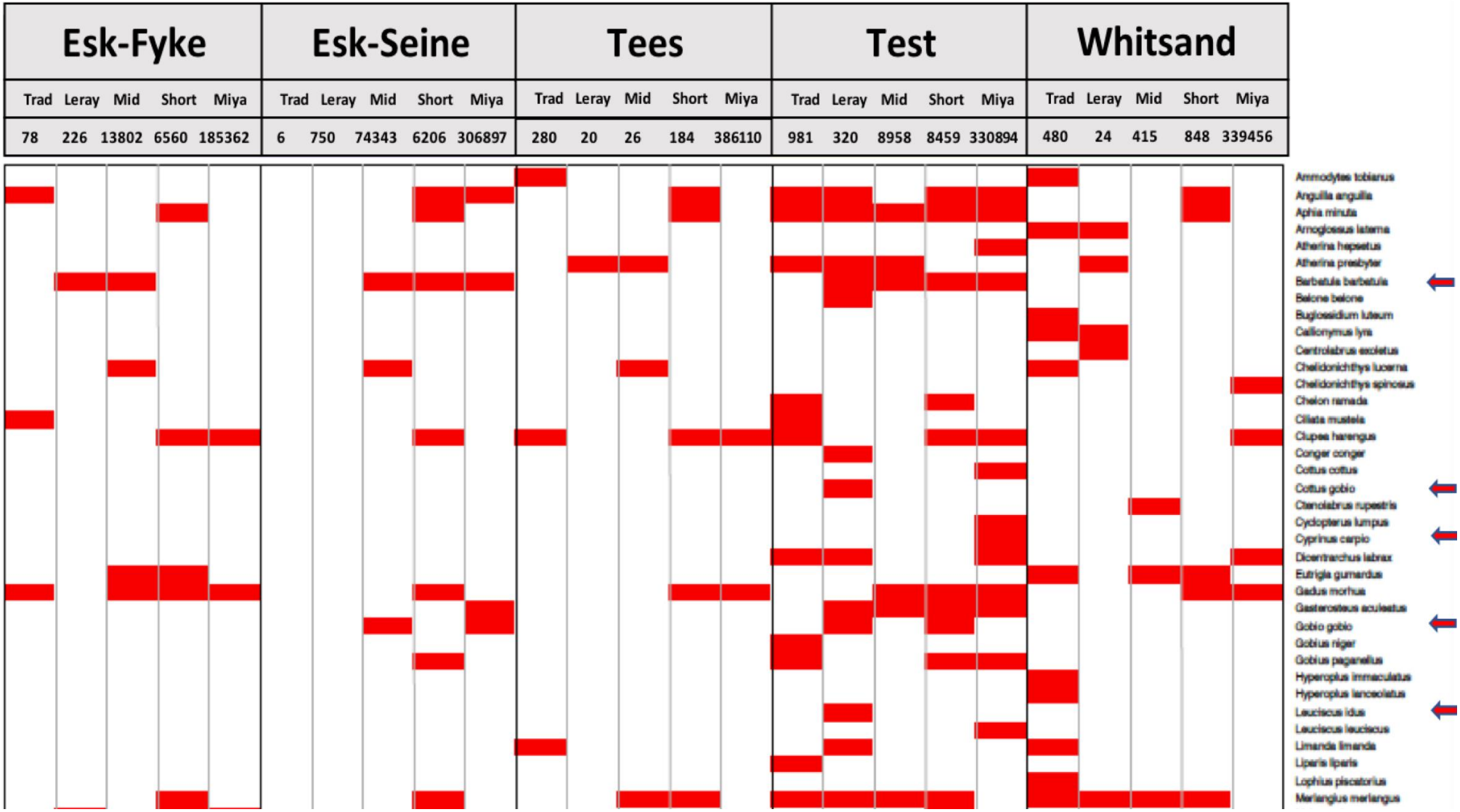


Figure 5.4 Venn diagram showing the distinctive and overlapping species detection between all four primer sets. Red stars indicating MOTUs taxonomically assigned to Pacific species and green stars indicating MOTUs assigned to Mediterranean species. The blue star indicates the invasive topmouth gudgeon (*Pseudorasbora parva*).

Figure 5.5 shows all the species detected by both traditional morphological sampling, and the eDNA surveys, for the four different primers, indicated per sampling site. The site with the highest density of species diversity, detected by both morphological and eDNA surveys, is the Test estuary. Here 22 species were identified by morphological surveys, and 44 species by eDNA metabarcoding. At every site, more teleost species were detected by the combined four primer sets than with the morphological surveys (Table 5.4.) However, none of the four primer sets individually outperforms the morphological sampling at all sites (only at 1,2 or 3 out of the 5 sites). Furthermore, the detection patterns overall are very irregular, and the matches between the morphological surveys and metabarcoding results are not very robust, indicating that a larger number of replicates would be needed (see also Figure 5.3). Ten

freshwater species were detected by eDNA but not the morphological surveys; *Barbatula barbatula* (stone loach), *Cottus gobio* (European bullhead), *Cyprinus carpio* (European carp), *Gobio gobio* (gudgeon), *Leuciscus idus* (ide), *Perca fluviatilis* (European perch), *Phoxinus phoxinus* (Eurasian minnow), *Pseudorasbora parva* (topmouth gudgeon), *Scardinius erythrophthalmus* (common rudd), and *Squalius cephalus* (Chub). It is most likely that the DNA from these species was transported down to the sampling estuaries from the upstream rivers (Deiner et al. 2016).


Esk-Fyke
Esk-Seine
Tees
Test
Whitsand
Trad Leray Mid Short Miya
78 226 13802 6560 185362
Trad Leray Mid Short Miya
6 750 74343 6206 306897
Trad Leray Mid Short Miya
280 20 26 184 386110
Trad Leray Mid Short Miya
981 320 8958 8459 330894
Trad Leray Mid Short Miya
480 24 415 848 339456
Ammodytes tobianus
Anguilla anguilla
Aphia minuta
Arnoglossus laterna
Atherina hepsetus
Atherina presbyter
Barbatula barbatula
Belone belone
Buglossidium luteum
Callionymus lyra
Centrolabrus exoletus
Chelidonichthys lucerna
Chelidonichthys spinosus
Chelon ramada
Ciliata mustela
Clupea harengus
Conger conger
Cottus cottus
Cottus gobio
Ctenolabrus rupestris
Cyclopterus lumpus
Cyprinus carpio
Dicentrarchus labrax
Eutrigla gurnardus
Gadus morhua
Gasterosteus aculeatus
Gobio gobio
Gobius niger
Gobius paganellus
Hyperoplus immaculatus
Hyperoplus lanceolatus
Leuciscus idus
Leuciscus leuciscus
Limanda limanda
Liparis liparis
Lophius piscatorius
Merlangius merlangus


Mohra mohra
Mullus surmuletus
Oncorhynchus mykiss
Pagrus pagrus
Pagusa lascaris
Perca fluviatilis
Phoxinus phoxinus
Platichthys flesus
Pleuronectes platessa
Pleuronectes quadrituberculatus
Pollachius pollachius
Pollachius virens
Pomatoschistus CBM_DNA_2008
Pomatoschistus knerii
Pomatoschistus minutus
Pomatoschistus pictus
Pseudorasbora parva
Rutilus rutilus
Salmo salar
Salmo trutta
Sardina pilchardus
Scardinius erythrophthalmus
Scomber scombrus
Scophthalmus maximus
Scophthalmus rhombus
Solea solea
Spondyliosoma cantharus
Sprattus sprattus
Squalius cephalus
Symphodus bailloni
Symphodus melops
Symphodus ocellatus
Syngnathus rostellatus
Syngnathus typhle
Taurulus bubalis
Thymallus thymallus
Trachinus vipera
Trachurus trachurus
Trisopterus luscus
Trisopterus minutus
Zaugopterus punctatus
Zeus faber
Zoarces viviparus

Figure 5.5 Presence/absence diagram showing all the species (in alphabetical order) detected in the five sampling sites. For every site, species detection is shown for traditional sampling ('Trad'; total number of individuals detected), and for the four different primers, for which the total number of sequences is shown. Indicated by the red arrows are the ten fresh water species

Table 5.4 Numbers of teleost species detected at each of the sampling sites, by both morphological and eDNA surveys. Number of species for eDNA surveys are given for the four primers combined and for each primer set individually. Coloured numbers indicate eDNA species detection performance compared to morphological surveys. Green indicates same or higher number of species detected and red indicates less species detected compared to morphological surveys.

Site	Number of species					
	Morphological sampling	eDNA sampling primers combined	Leray - XT	SeaDNA-Mid	SeaDNA-Short	Miya-MiFish
Esk –Fyke	**8**	**23**	**9**	**4**	**12**	**8**
Esk –Seine	**4**	**22**	**7**	**5**	**11**	**6**
Tees	**7**	**14**	**5**	**4**	**5**	**4**
Test	**22**	**44**	**23**	**9**	**16**	**25**
Whitsand	**21**	**27**	**10**	**7**	**9**	**9**

5.4.3 Teleost diversity and read abundance patterns

Violin plots of MOTU richness for each primer set (Fig. 5.6), show how the different sample values are distributed, by comparing the variable sample size distribution across the five different sites. The distribution of density (number of MOTUs per sample) is represented by the width of the plots. Even though there are pronounced differences in species detection between the primers (Fig. 5.4), and additionally, in the numbers of different MOTUs detected (Table 5.3), the patterns of MOTU richness between the four primers is strikingly similar. The first thing that becomes apparent is that when using any of the four primer sets, the Test estuary is the location with the highest MOTU (species) diversity. With a maximum of >20 MOTUS in one sample with both MiFish and SeaDNA-Short primers, and a minimum of 9 MOTUs per sample (for SeaDNA-Mid). While the Tees estuary shows the lowest MOTU richness per sample (between 2 and 7 MOTUs), with any of the four primer sets. MOTU richness is most diverse among the different samples of the Esk-Fyke site, as indicated by the size and

shape of its plots. The sites can be ordered from highest to lowest MOTU richness; (1) Test, (2) Esk-Fyke & Whitsand, (3) Esk-Seine, and (4) Tees. No samples contain less than 3 teleost MOTUs.

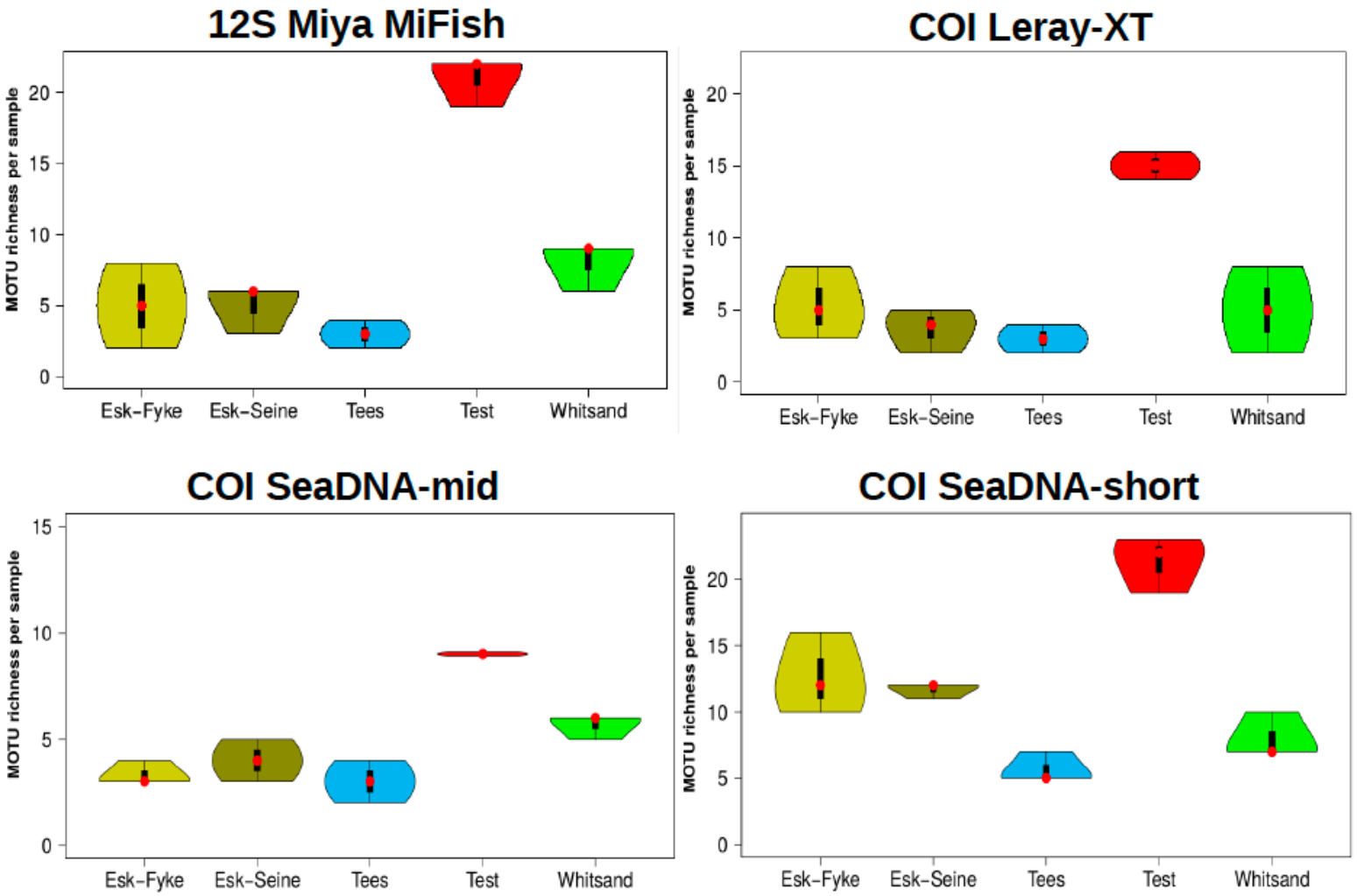


Figure 5.6 Violin plots showing detected teleost diversity (MOTU richness), per sample in the different sites by the four primers sets. The shapes indicate the density distribution of the samples, extending from the minimum to the maximum observed values. The median values are indicated by the red dots. The thick black bars are the interquartile ranges. The thin black extending lines represent the 95% confidence intervals such that the values in the wider parts of the plots are more probable than those in the narrower parts.

5.4.4 **Patterns of β-diversity**

Patterns of community structure between the five sampling sites are presented for all four primer sets in Figure 5.7, visualizing the dissimilarities among the different sites. Even though the species detection patterns between the four primers, within each site, are irregular (Fig. 5.5), it is apparent that samples cluster according to sampling site, with all four primers. The samples from Whitsand are most dissimilar from the other sites (with all primers), which is in accordance with the fact that this is a continental shelf sampling location, while the others are estuarine.

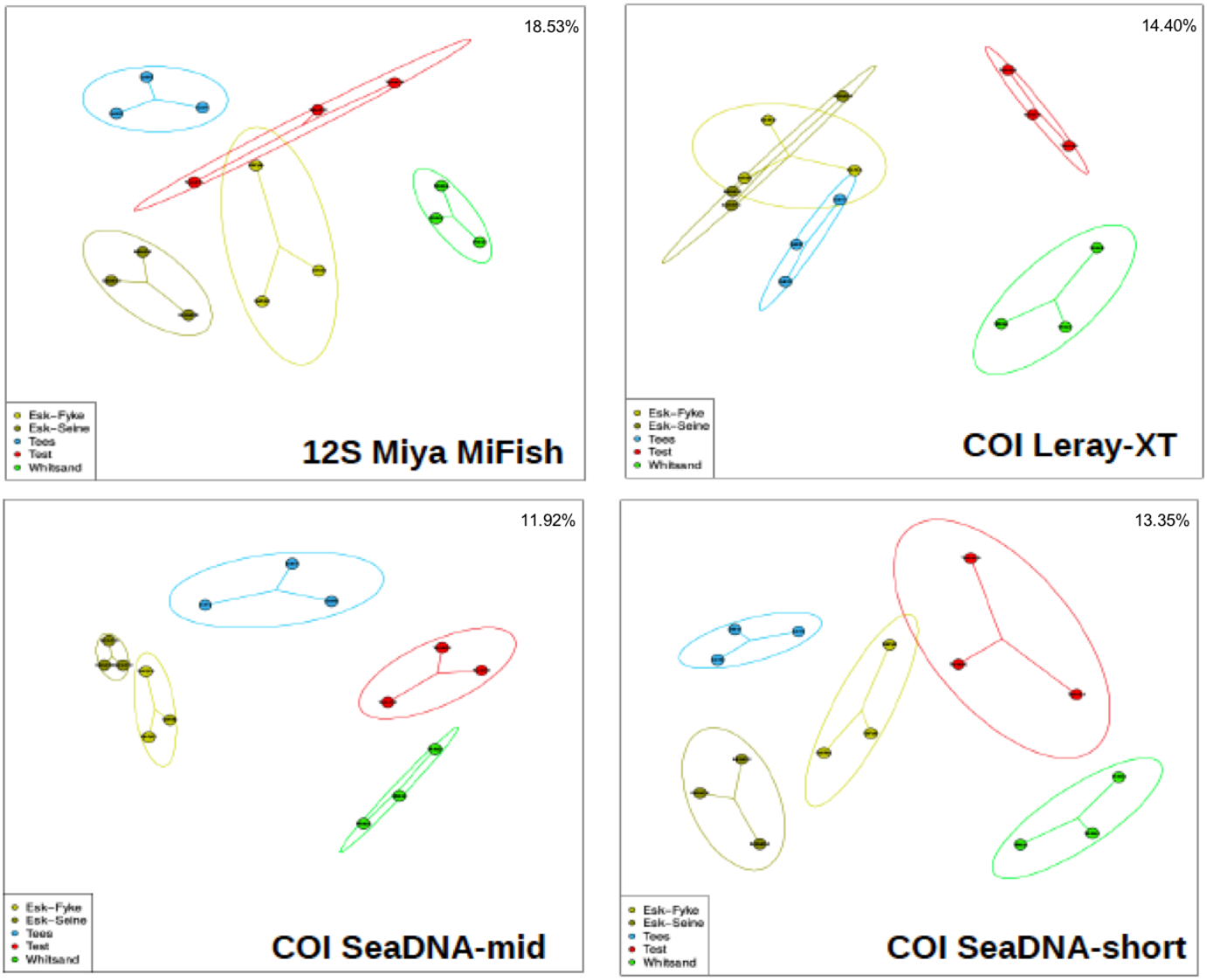


Figure 5.7 nMDS ordinations of the three replicate samples per location, as produced by each of the four primer sets. Results are based on the Jaccard index, based on the presence-absence data of MOTUs. Numbers in the upper right corners indicate stress of the final configurations

5.5 Discussion

The lack of suitable universal primers currently hinders the application of eDNA metabarcoding techniques for the characterization of teleost communities in estuarine and marine waters. Therefore, the goal of this study was to evaluate and compare the theoretical and practical performances of four primer sets, in amplifying a wide array of teleost diversity present in UK estuarine and marine ecosystems. Both *in silico*, and *in vitro* analysis allowed us to evaluate and compare the efficiency of two established (COI and 12S), and two novel (both COI) eDNA metabarcoding primer sets, in describing teleost diversity in five UK sampling sites, and additionally, to compare these results with concurrent morphological sampling.

As was predicted by *in silico* analysis (Table 5.2), the 12S MiFish primers have much greater specificity (77.7% of all reads was assigned to teleosts), compared to the three COI primers (Table 5.3). Strikingly, while with the Leray-XT primers (which are not designed specifically for the amplification of teleosts, but rather for eukaryotes in general) teleost read abundance only accounted for 0.05% of all reads (very low primer specificity), the number of species detected (MOTU richness) was nearly identical to that of the MiFish primers (32 and 33 respectively). However, there is an overlap of only 17 species between these two primers (Fig. 5.4), indicating that both are underperforming in species detection. Both primer sets show potential for applied eDNA metabarcoding of teleost diversity assessment in UK waters, however, significant improvements are required. For the 12S MiFish primers, that already have very high teleost specificity, it is mostly the reference database that is currently insufficient, particularly for the identification of Atlantic teleost species, as was indicated by the *in silico* analysis (only 51.9% of the 160 species was present in the reference database, Table 5.2), and highlighted by the taxonomic assignment of eDNA reads from UK sampling sites, to species that only occur in the Pacific Ocean or in the Mediterranean Sea (Fig. 5.4). However, irrespective of database completeness, some congeneric taxa will have exactly the same sequence within the relatively short length of the amplicon (165 bp), and hence, it will not be possible to distinguish those closely related species. While the COI Leray-XT markers allow for a high level of accurate species identification (taxonomic resolution of 98.6%, Table 5.2), their selectivity for teleost DNA is very low. Improving primer selectivity, while at the same time keeping broad taxonomic coverage for teleosts is very challenging, and it is this same challenge that lays at the base of the design of metabarcoding primers for any taxonomic group.

Even though the two other COI primers, SeaDNA-Mid and SeaDNA short, were designed specifically for this study, with the aim of improving selectivity for teleost sequences, they clearly performed less well compared with the MiFish and Leray-XT primer sets. Both of these primers were expected to perform at least equally well as the Leray-XT primers, based on the premises of producing a significantly shorter amplicon (130 and 55 bp respectively, compared to 313 bp for the Leray-XT primers), increasing the chances of amplifying the full length of the marker, from the degraded small fragments of DNA present in the water and enhancing the selectivity for teleost sequences. However, in order to improve taxonomic coverage of teleost sequences by keeping low levels of primer bias, both SeaDNA primers contain considerable levels of ambiguity (Table 5.1) in order to improve their universality (Wangensteen, Palacín, Guardiola, 2017), which in combination with the shorter amplicon length, was expected to result in better teleost detection performance. It has been suggested that 50-55bp may be the optimal amplicon length to acquire maximum resolution while concurrently optimizing (degraded) eDNA amplification by using a minimum amplicon length (Hajibabaei *et al.*, 2007; Wangensteen, Palacín, Guardiola, 2017). However, targeting short fragments may increase the probability of amplifying more of the eDNA present in a sample, it will also inevitably decrease taxonomic resolution of the marker, due to increased synonymies between species within these short sequences, rendering these primers less discriminant at the species level. Consequently, a large part of the teleost reads from both SeaDNA primers, could only be taxonomically assigned to genus or family level. Additionally, while for both SeaDNA primers, the forward primer is identical, the reverse primer of SeaDNA-Mid is prone to an increased level of primer bias due to a higher level of mismatches in the reverse primer binding site, resulting in a smaller taxonomic coverage (table 5.2), leading to a decreased level of taxonomic resolution compared to SeaDNA-Short. Accordingly, 17 and 18 teleost MOTUs were identified by SeaDNA-Mid and SeaDNA-Short respectively, but in each sampling site, SeaDNA-Short detected more species compared to SeaDNA-Mid (Table 5.4). Thus, our efforts to achieve an increase in the number of teleost reads while keeping high coverage levels and low primer bias, using COI-based markers, were largely unsuccessful.

Fish monitoring in the sampled range of transitional water bodies cannot be consistently applied, consequently, the catch data at each site is influenced by the different sampling approaches. However, the purpose of this study was firstly, to compare the performance of the different primers sets among each other and secondly, to investigate how eDNA metabarcoding compares to the standard traditional fish sampling, implemented in the

different locations, irrespective of sampling method. In spite of the fact that with a combination of the four primer sets, eDNA species detection was superior at every sampling site, and overall, 19 more species were detected with eDNA metabarcoding compared to the morphological surveys, the detection patterns between the two methods are quite dissimilar and 18 species present in the morphological samples have not been detected by eDNA (Figure 5.3). We expect that this issue will largely be resolved with increased eDNA sample sizes.

Our study reiterates that eDNA metabarcoding can be used to assess teleost diversity from water samples, however, even when a single taxonomic group is targeted, the use of different primer sets may still produce biased results in species detection within the same sample, indicating that the application of metabarcoding approaches for the detection of species-level diversity, using a single marker, still faces significant challenges. Thus, in order for eDNA metabarcoding to develop into a tool that can be used to assist fisheries professionals in the assessment and monitoring of fish communities, it is expected that in future applications for the assessment of community diversity for such taxonomically diverse groups as teleost, a multi-marker combination of primer sets may be most suitable to reduce taxonomic biases and increase taxonomic coverage and species detection probability, and hence provide a more accurate picture of species diversity.

5.6 Acknowledgements

For assistance with logistics, eDNA sampling and providing the morphological data, we thank Tony Gray and the crew of the Environment Agency at Newcastle Upon Tyne, Robin Somes and Peter Henderson from PISCES Conservation Ltd., and Aisling Smith and Sophie Rainbird from the Marine Biological Association of the UK in Plymouth. This work was co-funded by the Natural Environment Research Council, grant NE/N005759/1 (project SeaDNA), and the University of Salford R&E strategy funding,

6 Chapter VI

General Conclusion

The overarching theme of this thesis has been the development and application of an environmental DNA approach for the assessment of marine communities, with a special emphasis on elasmobranchs. The present body of work underpins the usefulness and potential implications of eDNA for large scale marine community assessment and monitoring. Environmental DNA is a tool with a growing list of research and conservation applications, in a wide range of environments. While eDNA analysis, applied to macro-organisms, was initially only used to gather presence-absence and distribution data of one or a few target species, other promising lines of enquiry now include population diversity assessment, estimates of abundances, population genetics, whole (mito)genome sequencing, diet characterisation and the description of trophic interactions. However, in order appropriately apply, and interpret the results of eDNA analysis, further research and method development is essential. Indeed, great effort will need to be directed towards primer design and the improvement of reference databases for a range of barcoding markers. Additionally, in order to gain a more comprehensive understanding of the ecology of eDNA and to increase the confidence with which interpret eDNA data sets may be interpreted, further investigations assessing how different environmental and taxa specific factors affect the production, degradation, distribution and detection of eDNA, is critical to both its current and future applications. This may be particularly true for elasmobranchs, as this group's unique biological and physiological traits are likely to influence eDNA production and detection.

As advances in molecular ecology, bioinformatics, and sequencing technologies continue to accelerate, eDNA methods have the potential to open up great possibilities for the applications of biodiversity assessments in general, but also specifically for the assessment and monitoring of shark communities. With a growing number of endangered and critically endangered shark and ray species, there is a clear urgency for a cost-effective, non-invasive and reliable method to obtain basic distribution, abundance and diversity data to advance conservation and management efforts. Many shark populations could benefit from the application of eDNA methods to study their diversity, abundance and distribution. Environmental DNA clearly has great potential to aid in conservation, remediation and restoration efforts of sharks and their relatives, as eDNA sampling can access inhospitable

environments, target elusive species, and provide a vast reduction in labour costs. In the future, it may be possible to implement mechanical sampling of (shark) eDNA. However, with over 95% of elasmobranchs found in marine ecosystems, and marine eDNA research still being in its infancy, compared to fresh water studies, it is important that research will focus on filling important knowledge gaps in order to improve the reliability of eDNA studies in marine ecosystems.

Conservation and management of marine biodiversity relies on the effective monitoring of species across large oceanic areas, where direct observation and identification of both large and minute individuals, are often complicated. While improvements to the method continue to be implemented, eDNA metabarcoding has great potential for developing into an effective assessment tool for marine biodiversity assessment, applicable to a wide range of ecological goals, from the detection of elusive species or populations, the mapping of diversity gradients in response to environmental variation, to the monitoring of the effectiveness of spatial protection measures. Environmental DNA applications have the potential to drastically enhance our ability to assess and monitor marine biodiversity, and shark presence and diversity in particular, which may lead the way to improved conservation strategies.

7 Appendix

Contamination control

One of the main challenges associated with the use of eDNA, is dealing with false positive and false negative detections. Due to the high sensitivity of eDNA methods, the most serious stumbling block is the risk of contamination, and hence the possibility of introducing false positive results. Contamination of samples may occur anywhere from preparing sampling equipment and collecting the samples in the field (target DNA being carried unintentionally from one locality to another), to every subsequent step of sample preparation, extraction and analysis in the laboratory. Hence, strict adherence to contamination control was followed at all field and laboratory stages in order to prevent the occurrence of contamination. The protocol used to minimize/avoid contamination in the field and in the lab including the use of disposable gloves (which are chanced when handling a new sample), and single use-sterile collection bottles and filtration equipment, and the bleaching (50% bleach) of sampling devices and all laboratory equipment and surfaces. Moreover, in the field, after every sampling event, the sampling devise was rinsed with tap water and left to dry out in the sun.. Additionally, a dedicated controlled eDNA lab at the University of Salford, with separate rooms designated for the physical separation of eDNA extraction, pre-PCR preparations and post-PCR procedures, was used for all laboratory work. During sample extraction, all laboratory equipment and surfaces were bleached after each sample treatment, prior to commencing with the next sample. Moreover, to identify potential contamination, DNA extraction blanks (elution buffer from extraction kit) and PCR blanks were included in all the library preparations and sequencing runs.

Supplementary materials

Sample	Code	Latitude	Longitude	Date	Raw reads	Reads after length filter	Elasmobranch reads	% Elasmo reads per total reads
Belize 1	BE1	16.742425	-87.826622	22/01/2015	29046	110	0	0.0000
Belize 2	BE2	16.739591	-87.825214	22/01/2015	27222	901	3	0.3330
Belize 3	BE3	16.753905	-87.829629	21/01/2015	16207	2504	0	0.0000
Belize 4	BE4	16.734768	-87.821916	21/01/2015	90688	3645	0	0.0000
Belize 5	BE5	16.734019	-87.823741	23/01/2015	4609	15	0	0.0000
Belize 6	BE6	16.720795	-87.873390	20/01/2015	209	22	1	4.5455
Belize 7	BE7	16.719798	-87.872900	24/01/2015	18341	572	0	0.0000
Belize 8	BE8	16.715376	-87.875989	20/01/2015	519772	137518	1	0.0007
Jamaica 1	JA1	18.470843	-77.402583	30/01/2015	39763	6549	0	0.0000
Jamaica 2	JA2	18.471855	-77.401480	02/02/2015	271092	11878	2	0.0168
Jamaica 3	JA3	18.462589	-77.407249	30/01/2015	75	4	0	0.0000
Jamaica 4	JA4	18.463221	-77.405946	02/02/2015	32474	561	3	0.5348
Jamaica 5	JA5	18.475094	-77.419606	30/01/2015	3851	622	0	0.0000
Jamaica 6	JA6	18.473628	-77.420228	02/02/2015	135780	2987	0	0.0000
Jamaica 7	JA7	18.509661	-77.918802	29/01/2015	527	6	2	33.3333
Jamaica 8	JA8	18.478746	-77.942029	28/01/2015	14643	51	1	1.9608
Jamaica 9	JA9	18.485105	-77.937299	29/01/2015	73	9	0	0.0000
Jamaica 10	JA10	18.493893	-77.935327	28/01/2015	30530	2005	0	0.0000
Jamaica 11	JA11	18.490476	-77.935679	29/01/2015	10093	71	1	1.4085
Turks&Caicos 1	TC1	21.489242	-71.544499	21/02/2015	8354	482	0	0.0000
Turks&Caicos 2	TC2	21.523266	-71.502067	24/02/2015	21106	369	1	0.2710
Turks&Caicos 3	TC3	21.491370	-71.517633	19/02/2015	5064	16	1	6.2500
Turks&Caicos 4	TC4	21.491240	-71.517264	23/02/2015	5662	539	0	0.0000
Turks&Caicos 5	TC5	21.524710	-71.514431	19/02/2015	115	12	0	0.0000
Turks&Caicos 6	TC6	21.524191	-71.515107	23/02/2015	24675	2775	0	0.0000
Turks&Caicos 7	TC7	21.509974	-71.546088	21/02/2015	64602	2694	10	0.3712
Turks&Caicos 8	TC8	21.507338	-71.549741	27/02/2015	105180	393	112	28.4987
Turks&Caicos 9	TC9	21.490505	-71.533662	21/02/2015	121	11	1	9.0909
Turks&Caicos 10	TC10	21.484253	-71.534393	20/02/2015	6167	614	1	0.1629
Turks&Caicos 11	TC11	21.483943	-71.534468	23/02/2015	59114	1935	0	0.0000
Turks&Caicos 12	TC12	21.453852	-71.568077	24/02/2015	354	42	6	14.2857
Turks&Caicos 13	TC13	21.454711	-71.569944	24/02/2015	76926	3851	8	0.2077
Turks&Caicos 14	TC14	21.454881	-71.572573	27/02/2015	81514	1596	0	0.0000
Turks&Caicos 15	TC15	21.448875	-71.567181	27/02/2015	102707	14445	4	0.0277
Turks&Caicos 16	TC16	21.484726	-71.520559	23/02/2015	131273	8674	0	0.0000
Turks&Caicos 17	TC17	21.483477	-71.519860	23/02/2015	4752	251	13	5.1793
Turks&Caicos 18	TC18	21.481869	-71.519097	24/02/2015	2385	194	4	2.0619
Turks&Caicos 19	TC19	21.484656	-71.521063	27/02/2015	183327	7253	6	0.0827
Turks&Caicos 20	TC20	21.456078	-71.563766	20/02/2015	17331	2902	15	0.5169
Bahamas 1	BH1	25.747522	-79.249553	09/02/2015	16482	733	2	0.2729
Bahamas 2	BH2	25.726562	-79.295653	07/02/2015	259	2	0	0.0000
Bahamas 3	BH3	25.726562	-79.295653	08/02/2015	1418	128	0	0.0000
Bahamas 4	BH4	25.726562	-79.295653	10/02/2015	65090	7592	48	0.6322
Bahamas 5	BH5	25.778076	-79.297283	09/02/2015	59751	1955	6	0.3069
Bahamas 6	BH6	25.702672	-79.303250	08/02/2015	48415	3781	744	19.6773
Bahamas 7	BH7	25.702940	-79.303027	09/02/2015	18033	55	2	3.6364
Bahamas 8	BH8	25.795269	-79.239509	09/02/2015	31744	560	4	0.7143
Bahamas 9	BH9	25.797046	-79.248568	13/02/2015	18462	1675	2	0.1194
Bahamas 10	BH10	25.748293	-79.262442	09/02/2015	3075	13	0	0.0000
Bahamas 11	BH11	25.677480	-79.284373	08/02/2015	3106	309	0	0.0000
Bahamas 12	BH12	25.651266	-79.293583	08/02/2015	17239	784	16	2.0408
Bahamas 13	BH13	25.651266	-79.293583	07/02/2015	1255	5	0	0.0000
Bahamas 14	BH14	25.64766	-79.312079	07/02/2015	10678	1842	553	30.0217
Bahamas 15	BH15	25.64766	-79.312079	07/02/2015	168701	3252	639	19.6494
Noumea 1	NO1	-22.37062	166.26869	26/11/2015	613	147	36	24.4898
Noumea 2	NO2	-22.37538	166.27646	26/11/2015	344	136	3	2.2059
Noumea 3	NO3	-22.38079	166.28214	26/11/2015	3320	125	5	4.0000
Noumea 4	NO4	-22.36531	166.25642	26/11/2015	6649	2201	59	2.6806
Noumea 5	NO5	-22.34095	166.23013	29/10/2015	6127	2871	173	6.0258
Noumea 6	NO6	-22.32607	166.22008	29/10/2015	13966	6	0	0.0000
New Caledonia North 1	NN1	-18.48525	163.0780833	06/09/2015	8556	790	505	63.9241
New Caledonia North 2	NN2	-18.45872	162.83798	07/09/2015	57480	2221	1449	65.2409
New Caledonia North 3	NN3	-18.49235	163.09102	17/11/2015	54559	10	3	30.0000
New Caledonia North 4	NN4	-19.47228	163.21939	12/11/2015	15468	3055	428	14.0098
New Caledonia North 5	NN5	-18.33913	162.9615	16/11/2015	2289	21	0	0.0000
New Caledonia North 6	NN6	-18.52115	162.87238	17/11/2015	5553	1101	2	0.1817
New Caledonia North 7	NN7	-19.35987	163.1638	11/11/2015	30418	3062	180	5.8785
New Caledonia North 8	NN8	-18.42886	162.99844	17/11/2015	7518	1141	408	35.7581
New Caledonia North 9	NN9	-18.50674	163.21638	15/11/2015	935	196	196	100.0000
New Caledonia North 10	NN10	-18.29657	163.00105	15/11/2015	67	3	2	66.6667
Chesterfield 1	NC1	-19.47889	158.25992	30/09/2015	4739	3987	3947	98.9967
Chesterfield 2	NC2	-19.21729	158.34457	30/09/2015	11239	1507	97	6.4366
Chesterfield 3	NC3	-19.06996	158.66454	04/09/2015	7006	5494	5205	94.7397
Chesterfield 4	NC4	-19.05768	158.98756	05/09/2015	43981	3055	89	2.9133
Chesterfield 5	NC5	-19.27486	158.97311	05/09/2015	22974	6062	4253	70.1584
Chesterfield 6	NC6	-19.42649	158.92831	05/09/2015	59599	5297	2290	43.2320
TOTAL N READS					**2972832**	**284252**	**21542**	
Average % elasmobranch reads per sample								12.1546

Supplementary Material 2.1 Sampling site information and total number of reads and number of elasmobranch reads per sample.

1. Paired-end alignment. Keep reads with quality > 40. Demultiplexing.

```
illuminapairedend -r SHAK_S1_L001_R2_001.fastq SHAK_S1_L001_R1_001.fastq | obiannotate -S goodali:'"Good_SHAK" if score>40.00 else "Bad_SHAK"' | obisplit -t goodali
ngsfilter -t ngsfilter_SHAK_fields.tsv --fasta-output -u unidentified_SHAK.fasta Good_SHAK.fasta > SHAK.filtered.fasta
```

2. Filter sequences with lengths between 120 and 135 bp and with only 'ACGT'.

```
obigrep -p 'seq_length>120' -p 'seq_length<135' -s '^[ACGT]+$' SHAK.filtered.fasta > SHAK.filtered_length.fasta
```

3. Group unique seqs.

```
obiuniq -m sample SHAK.filtered_length.fasta >  SHAK.unique.fasta
```

4. Change ids to a short index. Change format to vsearch. Remove chimeras.

```
obiannotate --seq-rank SHAK.unique.fasta | obiannotate --set-identifier '"'SHAK'%09d" % seq_rank' > SHAK.new.fasta
owi_obifasta2vsearch -i SHAK.new.fasta -o SHAK.vsearch.fasta
vsearch --uchime_denovo SHAK.vsearch.fasta --sizeout --nonchimeras SHAK.nonchimeras.fasta --chimeras SHAK.chimeras.fasta --uchimeout SHAK.uchimeout.txt
```

5. Cluster at 99% with sumaclust. Get cluster centers.

```
sumaclust -t 0.99 -s count -p 10 SHAK.nonchimeras.fasta > SHAK.sumaclust99.fasta
obigrep -p 'cluster_center' SHAK.sumaclust99.fasta > SHAK.sumaclust99.centers.fasta
```

6. Taxonomic assignment using ecotag.

```
ecotag -d taxo_sharks -R db_Elasmobranchii_Bakker_et_al_2017.fasta SHAK.sumaclust99.centers.fasta > SHAK.ecotag.fasta
```

7. Add taxa above order level.

```
owi_add_taxonomy -i SHAK.ecotag.fasta -o SHAK.ecotag.fasta.annotated.csv
```

8. Recount abundances by sample.

```
obitab -o SHAK.sumaclust99.fasta >  SHAK.sumaclust99.tab
owi_recount_sumaclust          -i          SHAK.sumaclust99.tab          -o
SHAK.sumaclust99.counts.csv
```

9. Combine ecotag and abundance files.

```
owi_combine          -i          SHAK.ecotag.fasta.annotated.csv          -a
SHAK.sumaclust99.counts.csv -o SHAK_all_MOTUs.csv
```

10. Collapse MOTUs.

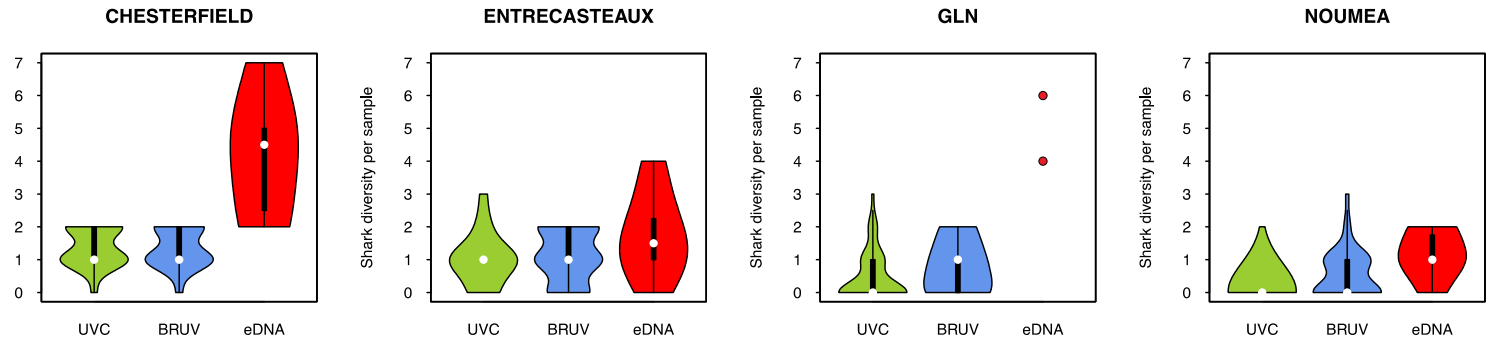
```
owi_collapse -s 13 -e 88 -i SHAK_all_MOTUs.csv
```

11. Curate the dataset manually.

12. Re-collapse MOTUs after curating.

```
owi_collapse -s 13 -e 88 -i SHAK_all_MOTUs_curated.csv
```

Supplementary material 2.3 and 3.3 Metabarcoding pipeline for COI Elasmobranchii Fields et al. primers



Supplementary material 3.1 Number of shark species per sample in overlapping collection sites. Violin plot showing detected shark species richness by the different methods in Chesterfield, Entrecasteaux, GLN and Nouméa. Only two eDNA samples were collected in GLN (red dots). White dots are mean values; thick black bars correspond to interquartile ranges; thin black lines are 95% confidence intervals.

Reverse primers		Forward primers 1		Forward primers 2	
Shark-COI-MINIR_tag01	NNNNaacaagccAAGATTACAAAAGCGTGGGC	**FishF1_tag01**	NNaacaagccTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag01**	NNaacaagccTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag02	NNNggaatgagAAGATTACAAAAGCGTGGGC	**FishF1_tag02**	NNNggaatgagTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag02**	NNNggaatgagTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag03	NNaattgccgAAGATTACAAAAGCGTGGGC	**FishF1_tag03**	NNNNaattgccgTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag03**	NNNNaattgccgTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag04	NNNNcgaccataAAGATTACAAAAGCGTGGGC	**FishF1_tag04**	NNcgaccataTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag04**	NNcgaccataTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag05	NNNatgctgacAAGATTACAAAAGCGTGGGC	**FishF1_tag05**	NNNatgctgacTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag05**	NNNatgctgacTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag06	NNtgagacagAAGATTACAAAAGCGTGGGC	**FishF1_tag06**	NNNNtgagacagTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag06**	NNNNtgagacagTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag07	NNNNgagcttacAAGATTACAAAAGCGTGGGC	**FishF1_tag07**	NNgagcttacTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag07**	NNgagcttacTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag08	NNNttaccaggAAGATTACAAAAGCGTGGGC	**FishF1_tag08**	NNNttaccaggTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag08**	NNNttaccaggTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag09	NNtgagagctAAGATTACAAAAGCGTGGGC	**FishF1_tag09**	NNNNtgagagctTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag09**	NNNNtgagagctTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag10	NNNNctgaccttAAGATTACAAAAGCGTGGGC	**FishF1_tag10**	NNctgaccttTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag10**	NNctgaccttTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag11	NNNatgcttggAAGATTACAAAAGCGTGGGC	**FishF1_tag11**	NNNatgcttggTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag11**	NNNatgcttggTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag12	NNaacaccgtAAGATTACAAAAGCGTGGGC	**FishF1_tag12**	NNNNaacaccgtTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag12**	NNNNaacaccgtTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag13	NNNNttaccgctAAGATTACAAAAGCGTGGGC	**FishF1_tag13**	NNttaccgctTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag13**	NNttaccgctTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag14	NNNccagtatgAAGATTACAAAAGCGTGGGC	**FishF1_tag14**	NNNccagtatgTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag14**	NNNccagtatgTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag15	NNtgagatgcAAGATTACAAAAGCGTGGGC	**FishF1_tag15**	NNNNtgagatgcTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag15**	NNNNtgagatgcTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag16	NNNNgtgcaactAAGATTACAAAAGCGTGGGC	**FishF1_tag16**	NNgtgcaactTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag16**	NNgtgcaactTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag17	NNNacaaccgaAAGATTACAAAAGCGTGGGC	**FishF1_tag17**	NNNacaaccgaTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag17**	NNNacaaccgaTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag18	NNtgagcctaAAGATTACAAAAGCGTGGGC	**FishF1_tag18**	NNNNtgagcctaTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag18**	NNNNtgagcctaTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag19	NNNNatggaggtAAGATTACAAAAGCGTGGGC	**FishF1_tag19**	NNatggaggtTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag19**	NNatggaggtTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag20	NNNtcatacgcAAGATTACAAAAGCGTGGGC	**FishF1_tag20**	NNNtcatacgcTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag20**	NNNtcatacgcTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag21	NNctgagtctAAGATTACAAAAGCGTGGGC	**FishF1_tag21**	NNNNctgagtctTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag21**	NNNNctgagtctTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag22	NNNNgaggtgaaAAGATTACAAAAGCGTGGGC	**FishF1_tag22**	NNgaggtgaaTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag22**	NNgaggtgaaTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag23	NNNggcatgtaAAGATTACAAAAGCGTGGGC	**FishF1_tag23**	NNNggcatgtaTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag23**	NNNggcatgtaTCGACTAATCATAAAGATATCGGCAC
Shark-COI-MINIR_tag24	NNgtgccataAAGATTACAAAAGCGTGGGC	**FishF1_tag24**	NNNNgtgccataTCAACCAACCACAAAGACATTGGCAC	**FishF2_tag24**	NNNNgtgccataTCGACTAATCATAAAGATATCGGCAC

Supplementary material 3.2 Full sequences of the 24 tagged primer sets used. The primer mix for each PCR included the reverse primer and an equimolar mixture of the two forward primers, all tagged with the same 8-bp tag (in lowercase in this table). A variable number of fully degenerate positions (Ns) was added at the beginning of each primer, to increase sequence diversity

Sample	Details	Latitude	Longitude	Date
Belize 1	Glovers Back Reef	16.742425	-87.826622	22/01/2015
Belize 2	Glovers Back Reef	16.739591	-87.825214	22/01/2015
Belize 3	Glovers Back Reef	16.753905	-87.829629	21/01/2015
Belize 4	Glovers Fore Reef	16.734768	-87.821916	21/01/2015
Belize 5	Glovers Fore Reef	16.734019	-87.823741	23/01/2015
Belize 6	S-W Key Back Reef	16.720795	-87.873390	20/01/2015
Belize 7	S-W Key Back Reef	16.719798	-87.872900	24/01/2015
Belize 8	S-W Key Fore Reef	16.715376	-87.875989	20/01/2015
Jamaica 1	Discovery Bay	18.470843	-77.402583	30/01/2015
Jamaica 2	Discovery Bay	18.471855	-77.401480	02/02/2015
Jamaica 3	Discovery Bay	18.462589	-77.407249	30/01/2015
Jamaica 4	Discovery Bay	18.463221	-77.405946	02/02/2015
Jamaica 5	Discovery Bay	18.475094	-77.419606	30/01/2015
Jamaica 6	Discovery Bay	18.473628	-77.420228	02/02/2015
Jamaica 7	Montego Bay	18.509661	-77.918802	29/01/2015
Jamaica 8	Montego Bay	18.478746	-77.942029	28/01/2015
Jamaica 9	Montego Bay	18.485105	-77.937299	29/01/2015
Jamaica 10	Montego Bay	18.493893	-77.935327	28/01/2015
Jamaica 11	Montego Bay	18.490476	-77.935679	29/01/2015
Jamaica 12	Montego Bay	18.509661	-77.918802	29/01/2015
Turks&Caicos 1	Reef	21.489242	-71.544499	21/02/2015
Turks&Caicos 2	Mangroves	21.523266	-71.502067	24/02/2015
Turks&Caicos 3	Mangroves	21.491370	-71.517633	19/02/2015
Turks&Caicos 4	Mangroves	21.491240	-71.517264	23/02/2015
Turks&Caicos 5	Mangroves	21.524710	-71.514431	19/02/2015
Turks&Caicos 6	Mangroves	21.524191	-71.515107	23/02/2015
Turks&Caicos 7	Reef	21.509974	-71.546088	21/02/2015
Turks&Caicos 8	Reef	21.507338	-71.549741	27/02/2015
Turks&Caicos 9	Marina	21.490505	-71.533662	21/02/2015
Turks&Caicos 10	Reef	21.484253	-71.534393	20/02/2015
Turks&Caicos 11	Reef	21.483943	-71.534468	23/02/2015
Turks&Caicos 12	Reef	21.453852	-71.568077	24/02/2015
Turks&Caicos 13	Reef	21.454711	-71.569944	24/02/2015
Turks&Caicos 14	Reef	21.454881	-71.572573	27/02/2015
Turks&Caicos 15	Reef	21.448875	-71.567181	27/02/2015
Turks&Caicos 16	Reef	21.484726	-71.520559	23/02/2015
Turks&Caicos 17	Reef	21.483477	-71.519860	23/02/2015
Turks&Caicos 18	Reef	21.481869	-71.519097	24/02/2015
Turks&Caicos 19	Reef	21.484656	-71.521063	27/02/2015
Turks&Caicos 20	Reef	21.456078	-71.563766	20/02/2015
Bahamas 1	Mangroves	25.747522	-79.249553	09/02/2015
Bahamas 2	Reef	25.726562	-79.29565	07/02/2015
Bahamas 3	Reef	25.726562	-79.29565	08/02/2015
Bahamas 4	Reef	25.726562	-79.29565	10/02/2015
Bahamas 5	Reef	25.778076	-79.29728	09/02/2015
Bahamas 6	Reef	25.702672	-79.303250	08/02/2015
Bahamas 7	Reef	25.702940	-79.303027	09/02/2015
Bahamas 8	Reef	25.795269	-79.239509	09/02/2015
Bahamas 9	Reef	25.797046	-79.248568	13/02/2015
Bahamas 10	Mangroves	25.748293	-79.26244	09/02/2015
Bahamas 11	Reef	25.677480	-79.284373	08/02/2015
Bahamas 12	Reef	25.651266	-79.29358	08/02/2015
Bahamas 13	Reef	25.651266	-79.29358	07/02/2015
Bahamas 14	Reef	25.64766	-79.31208	07/02/2015
Bahamas 15	Reef	25.64766	-79.31208	07/02/2015
British Virgin Islands 1	Shallow Shore	18.416065	-64.586189	25/02/2017
British Virgin Islands 2	Shallow Shore	18.423109	-64.613564	25/02/2017
British Virgin Islands 3	Shallow Shore	18.449133	-64.645255	25/02/2017
British Virgin Islands 4	Reef	18.385451	-64.471362	26/02/2017
British Virgin Islands 5	Reef	18.372929	-64.488524	26/02/2017
British Virgin Islands 6	Shallow Shore	18.453958	-64.575063	01/03/2017
British Virgin Islands 7	Bay	18.498384	-64.415018	27/02/2017
British Virgin Islands 8	Bay	18.498384	-64.415018	27/02/2017
British Virgin Islands 9	Bay	18.498384	-64.415018	27/02/2017
British Virgin Islands 10	Bay	18.498384	-64.415018	27/02/2017
British Virgin Islands 11	Open Reef	18.511643	-64.352053	27/02/2017
British Virgin Islands 12	Open Reef	18.511643	-64.352053	27/02/2017
British Virgin Islands 13	Open Reef	18.511643	-64.352053	27/02/2017

Supplementary material 4.1 List with sampling site information

id	rank	scientific_name	best_id	superkingdom	kingdom	phylum	av_Rest_Carib	av_Brit_Virgin_Islands	cumsum
SHOK_000000703	superkingdom	Eukaryota	0.7645				29.298	0.87	17.055
SHOK_000071024	species	Rhizosolenia setigera	0.8508	Chromalveolata	Stramenopiles	Bacillariophyta	0	6.524	20.952
SHOK_000000006	species	Micromonas commoda	0.9872	Archaeplastida	Viridiplantae	Chlorophyta	5.753	2.255	24.647
SHOK_000000218	superkingdom	Eukaryota	0.7688				2.101	5.203	27.125
SHOK_000000152	class	Dinophyceae	0.9778	Chromalveolata	Alveolata	Dinoflagellata	4.991	1.206	29.494
SHOK_000000095	superkingdom	Eukaryota	0.8255				3.59	0	31.639
SHOK_000006526	kingdom	Stramenopiles	0.8019	Chromalveolata	Stramenopiles		0.006	3.452	33.697
SHOK_000002291	class	Oomycetes	0.8297	Chromalveolata	Stramenopiles	Oomycota	0.833	3.218	35.304
SHOK_000001753	no rank	Micromonas pusilla CCMP1545	0.7631	Archaeplastida	Viridiplantae	Chlorophyta	0.716	2.276	36.465
SHOK_000017818	kingdom	Stramenopiles	0.8193	Chromalveolata	Stramenopiles		0.608	2.43	37.613
SHOK_000006338	kingdom	Metazoa	0.7844	Opisthokonta	Metazoa		0.112	1.601	38.502
SHOK_000009258	superkingdom	Eukaryota	0.7822				1.506	2.28	39.372
SHOK_000001063	kingdom	Metazoa	0.8224	Opisthokonta	Metazoa		0.045	1.39	40.177
SHOK_000113134	kingdom	Stramenopiles	0.8307	Chromalveolata	Stramenopiles		0.029	1.365	40.975
SHOK_002644370	order	Spionida	0.8202	Opisthokonta	Metazoa	Annelida	0	1.239	41.715
SHOK_000000001	superkingdom	Eukaryota	0.7563				1.171	0.019	42.404
SHOK_002643991	order	Spionida	0.8266	Opisthokonta	Metazoa	Annelida	0	1.098	43.061
SHOK_000005655	superkingdom	Eukaryota	0.7901				0.04	1.069	43.68
SHOK_000000932	superkingdom	Eukaryota	0.8261				0.354	1.221	44.286
SHOK_002644124	no rank	Eumetazoa	0.858	Opisthokonta	Metazoa		0	1.015	44.893
SHOK_000011956	superkingdom	Eukaryota	0.8276				0.095	1.007	45.451
SHOK_000000523	kingdom	Stramenopiles	0.8287	Chromalveolata	Stramenopiles		1.107	0.79	45.998
SHOK_000000898	kingdom	Stramenopiles	0.8025	Chromalveolata	Stramenopiles		1.017	0.134	46.542
SHOK_000000400	phylum	Bacillariophyta	0.854	Chromalveolata	Stramenopiles	Bacillariophyta	0.119	0.937	47.086
SHOK_000002122	superkingdom	Eukaryota	0.7652				0.905	0.003	47.625
SHOK_000001047	species	Hematodinium sp.	0.8148	Chromalveolata	Alveolata	Dinoflagellata	0.027	0.904	48.149
SHOK_000001896	superkingdom	Eukaryota	0.8318				0.15	0.904	48.637
SHOK_000000966	superkingdom	Eukaryota	0.7671				0.816	0.048	49.103
SHOK_000029478	kingdom	Metazoa	0.81	Opisthokonta	Metazoa		0.013	0.786	49.565
SHOK_000000398	phylum	Arthropoda	0.787	Opisthokonta	Metazoa	Arthropoda	0.145	0.739	50.011

Supplementary material 4.3 Results SIMPER analysis for all locations vs the British Virgin Islands. List of the 30 most abundant (in reads) MOTUs among the ones identified by SIMPER as contributing more to the differentiation between the analyzed locations. Id is the identifier of the MOTU in the dataset. Av are the mean abundances (in semiquantitative ranks) of the MOTUs in the different sites. Cumsum is cumulative percentage of the MOTUs contributing to the differentiation. Rank refers to the taxonomic rank that ecotage could assign to the MOTUs.

id	rank	scientific_name	best_id	superkingdom	kingdom	phylum	av_Mangrove	av_Reef	cumsum
SHOK_000000152	class	Dinophyceae	0.9778	Chromalveolata	Alveolata	Dinoflagellata	21.47	2.526	11.732
SHOK_000000703	superkingdom	Eukaryota	0.7645				9.518	21.338	20.851
SHOK_000000006	species	Micromonas commoda	0.9872	Archaeplastida	Viridiplantae	Chlorophyta	0	12.64	28.679
SHOK_000000095	superkingdom	Eukaryota	0.8255				0	10.462	35.158
SHOK_000000218	superkingdom	Eukaryota	0.7688				10.966	1.647	41.314
SHOK_000001282	superkingdom	Eukaryota	0.8025				7.218	0.06	45.747
SHOK_000006586	kingdom	Metazoa	0.7774	Opisthokonta	Metazoa		3.43	0.016	47.868
SHOK_000004628	genus	Heterocapsa	0.8129	Chromalveolata	Alveolata	Dinoflagellata	2.704	0.004	49.54
SHOK_000002122	superkingdom	Eukaryota	0.7652				0	2.632	51.17
SHOK_000001038	species	Nitzschia AO711	0.8825	Chromalveolata	Stramenopiles	Bacillariophyta	1.879	0.012	52.326
SHOK_000009258	superkingdom	Eukaryota	0.7822				0.221	1.739	53.267
SHOK_000012709	superkingdom	Eukaryota	0.7982				0.437	1.501	54.197
SHOK_000006708	superkingdom	Eukaryota	0.7423				0	1.425	55.079
SHOK_000018954	kingdom	Metazoa	0.7799	Opisthokonta	Metazoa		1.392	0.009	55.938
SHOK_000000375	species	Clausocalanus furcatus	0.984	Opisthokonta	Metazoa	Arthropoda	0	1.365	56.784
SHOK_000000898	kingdom	Stramenopiles	0.8025	Chromalveolata	Stramenopiles		0.956	1.952	57.539
SHOK_000005148	genus	Heterocapsa	0.8179	Chromalveolata	Alveolata	Dinoflagellata	1.172	0.034	58.243
SHOK_000000507	class	Mamiellophyceae	0.8062	Archaeplastida	Viridiplantae	Chlorophyta	0	1.105	58.927
SHOK_000015472	no rank	Protostomia	0.8176	Opisthokonta	Metazoa		1.08	0	59.596
SHOK_000000523	kingdom	Stramenopiles	0.8287	Chromalveolata	Stramenopiles		1.623	1.391	60.237
SHOK_000008247	no rank	Protostomia	0.7943	Opisthokonta	Metazoa		0.996	0.001	60.853
SHOK_000000433	superkingdom	Eukaryota	0.7584				0	0.973	61.455
SHOK_000000524	superkingdom	Eukaryota	0.7631				0	0.927	62.03
SHOK_000039982	suborder	Aplousobranchia	0.7886	Opisthokonta	Metazoa	Chordata	0.864	0.007	62.565
SHOK_000002291	class	Oomycetes	0.8297	Chromalveolata	Stramenopiles	Oomycota	0	0.851	63.092
SHOK_000013410	species	Gymnodinium impudicum	0.9747	Chromalveolata	Alveolata	Dinoflagellata	0.756	0.001	63.559
SHOK_000000786	superkingdom	Eukaryota	0.787				0.281	0.944	64.026
SHOK_000000357	class	Dinophyceae	0.9462	Chromalveolata	Alveolata	Dinoflagellata	0.8	0.053	64.489
SHOK_000023584	infraclass	Neocopepoda	0.8239	Opisthokonta	Metazoa	Arthropoda	0.719	0	64.934
SHOK_000000001	superkingdom	Eukaryota	0.7563				0.46	0.941	65.364

Supplementary material 4.3 Results SIMPER analysis for the mangrove vs reef sites in the Bahamas

id	rank	scientific_name	best_id	superkingdom	kingdom	phylum	av_Back_Reef	av_Fore_Reef	cumsum
SHOK_000000703	superkingdom	Eukaryota	0.7645				15.638	41.845	21.503
SHOK_000000152	class	Dinophyceae	0.9778	Chromalveolata	Alveolata	Dinoflagellata	6.166	1.526	25.397
SHOK_000000218	superkingdom	Eukaryota	0.7688				6.136	2.033	28.763
SHOK_000004295	order	Eunicida	0.8208	Opisthokonta	Metazoa	Annelida	4.008	0	32.052
SHOK_000000192	superkingdom	Eukaryota	0.7692				0	3.965	35.305
SHOK_000000523	kingdom	Stramenopiles	0.8287	Chromalveolata	Stramenopiles		3.167	0.51	37.485
SHOK_000009561	suborder	Aplousobranchia	0.8222	Opisthokonta	Metazoa	Chordata	2.495	0.011	39.523
SHOK_000002291	class	Oomycetes	0.8297	Chromalveolata	Stramenopiles	Oomycota	1.167	3.234	41.418
SHOK_000003958	order	Anthoathecata	0.8639	Opisthokonta	Metazoa	Cnidaria	0	2.23	43.248
SHOK_000000898	kingdom	Stramenopiles	0.8025	Chromalveolata	Stramenopiles		2.301	0.446	44.769
SHOK_000000197	superkingdom	Eukaryota	0.7938				0.298	2.094	46.243
SHOK_000015168	phylum	Annelida	0.8196	Opisthokonta	Metazoa	Annelida	1.633	0	47.582
SHOK_000011893	superkingdom	Opisthokonta	0.7944	Opisthokonta			0	1.632	48.922
SHOK_000009258	superkingdom	Eukaryota	0.7822				4.076	3.288	50.178
SHOK_000000347	no rank	Eumetazoa	0.8	Opisthokonta	Metazoa		1.502	0	51.41
SHOK_000001753	no rank	Micromonas pusilla CCMP1545	0.7631	Archaeplastida	Viridiplantae	Chlorophyta	2.214	0.941	52.455
SHOK_000013270	superkingdom	Eukaryota	0.8418				1.133	0.02	53.368
SHOK_000017818	kingdom	Stramenopiles	0.8193	Chromalveolata	Stramenopiles		2.31	1.573	54.228
SHOK_000000001	superkingdom	Eukaryota	0.7563				0.832	1.836	55.063
SHOK_000000400	phylum	Bacillariophyta	0.854	Chromalveolata	Stramenopiles	Bacillariophyta	1	0.013	55.881
SHOK_000001731	superkingdom	Eukaryota	0.7864				0.952	0	56.663
SHOK_000014311	species	Thalassionema nitzschioides	0.8694	Chromalveolata	Stramenopiles	Bacillariophyta	0.924	0	57.421
SHOK_000000914	superkingdom	Eukaryota	0.7317				0.328	0.986	58.176
SHOK_000001896	superkingdom	Eukaryota	0.8318				1.063	0.198	58.886
SHOK_000006586	kingdom	Metazoa	0.7774	Opisthokonta	Metazoa		0.83	0	59.567
SHOK_000000725	species	Pseudo-nitzschia lineola	0.9019	Chromalveolata	Stramenopiles	Bacillariophyta	0.825	0	60.244
SHOK_000001019	superkingdom	Eukaryota	0.7638				0.575	0.991	60.915
SHOK_000013576	kingdom	Metazoa	0.7554	Opisthokonta	Metazoa		0.771	1.516	61.546
SHOK_000000966	superkingdom	Eukaryota	0.7671				0.22	0.979	62.173
SHOK_000028083	kingdom	Metazoa	0.7771	Opisthokonta	Metazoa		0.757	0.004	62.792

Supplementary material 4.3 Results SIMPER analysis for the back-reef vs fore-reef sites in Belize

id	rank	scientific_name	best_id	superkingdom	kingdom	phylum	av_Reef	av_Shore	cumsum
SHOK_000000703	superkingdom	Eukaryota	0.7645				42.347	33.386	21.405
SHOK_000000152	class	Dinophyceae	0.9778	Chromalveolata	Alveolata	Dinoflagellata	6.319	6.753	27.118
SHOK_000000184	class	Bivalvia	0.8217	Opisthokonta	Metazoa	Mollusca	0	4.574	31.276
SHOK_000000486	superkingdom	Eukaryota	0.76				0.622	4.169	35.348
SHOK_000001753	no rank	Micromonas pusilla CCMP1545	0.7631	Archaeplastida	Viridiplantae	Chlorophyta	0.321	1.537	36.909
SHOK_000005316	no rank	Eumetazoa	0.7695	Opisthokonta	Metazoa		1.492	0	38.265
SHOK_000002814	no rank	Eumetazoa	0.8106	Opisthokonta	Metazoa		1.417	0	39.553
SHOK_000001303	no rank	Protostomia	0.8069	Opisthokonta	Metazoa		0	1.28	40.716
SHOK_000006314	superkingdom	Eukaryota	0.7778				0.17	1.052	41.773
SHOK_000029624	suborder	Aplousobranchia	0.7962	Opisthokonta	Metazoa	Chordata	1.016	0.312	42.774
SHOK_000000218	superkingdom	Eukaryota	0.7688				0.576	1.211	43.744
SHOK_000000230	no rank	Eumetazoa	0.8333	Opisthokonta	Metazoa		0.998	0	44.652
SHOK_000001012	species	Cunina fowleri	0.7981	Opisthokonta	Metazoa	Cnidaria	0.993	0	45.554
SHOK_000001190	superkingdom	Eukaryota	0.753				1.342	1.196	46.414
SHOK_000000279	superkingdom	Eukaryota	0.7477				0.138	0.766	47.173
SHOK_000000966	superkingdom	Eukaryota	0.7671				1.449	1.077	47.917
SHOK_000000001	superkingdom	Eukaryota	0.7563				1.546	1.215	48.66
SHOK_000000523	kingdom	Stramenopiles	0.8287	Chromalveolata	Stramenopiles		0.605	0.837	49.328
SHOK_000001742	superkingdom	Eukaryota	0.7695				1.048	0.764	49.994
SHOK_000002150	species	Heterocapsa circularisquama	0.8339	Chromalveolata	Alveolata	Dinoflagellata	0.708	0.02	50.65
SHOK_000000506	order	Terebellida	0.8182	Opisthokonta	Metazoa	Annelida	0.002	0.712	51.297
SHOK_000009258	superkingdom	Eukaryota	0.7822				0.826	0.634	51.915
SHOK_000001659	class	Dinophyceae	0.8851	Chromalveolata	Alveolata	Dinoflagellata	0.359	0.425	52.532
SHOK_000013576	kingdom	Metazoa	0.7554	Opisthokonta	Metazoa		1.064	0.845	53.143
SHOK_000000197	superkingdom	Eukaryota	0.7938				0.673	0.585	53.752
SHOK_000000898	kingdom	Stramenopiles	0.8025	Chromalveolata	Stramenopiles		0.439	0.754	54.349
SHOK_000006207	superkingdom	Eukaryota	0.716				0.666	0.089	54.942
SHOK_000000932	superkingdom	Eukaryota	0.8261				0.369	0.341	55.503
SHOK_000000848	superkingdom	Eukaryota	0.7576				1.135	0.922	56.056
SHOK_000028083	kingdom	Metazoa	0.7771	Opisthokonta	Metazoa		0.44	0.262	56.603

Supplementary material 4.3 Results SIMPER analysis for the reef vs shore sites in the British Virgin Islands

id	rank	scientific_name	best_id	superkingdom	kingdom	phylum	av_Discovery_Bay	av_Montego_Bay	cumsum
SHOK_000000703	superkingdom	Eukaryota	0.7645				28.265	30.501	8.767
SHOK_000000006	species	Micromonas commoda	0.9872	Archaeplastida	Viridiplantae	Chlorophyta	13.67	9.758	16.85
SHOK_000000095	superkingdom	Eukaryota	0.8255				4.705	0.96	21.069
SHOK_000002128	phylum	Annelida	0.817	Opisthokonta	Metazoa	Annelida	2.641	0.222	24.115
SHOK_000000152	class	Dinophyceae	0.9778	Chromalveolata	Alveolata	Dinoflagellata	4.331	1.797	27.14
SHOK_000000054	superkingdom	Eukaryota	0.7387				0.018	2.654	30.123
SHOK_000004383	class	Anthozoa	0.9553	Opisthokonta	Metazoa	Cnidaria	0	1.884	32.236
SHOK_000002814	no rank	Eumetazoa	0.8106	Opisthokonta	Metazoa		0	1.745	34.194
SHOK_000010093	species	Phascolosoma nigrescens	0.9744	Opisthokonta	Metazoa	Annelida	0.625	1.271	36.084
SHOK_000008415	order	Sabellida	0.8307	Opisthokonta	Metazoa	Annelida	0	1.628	37.91
SHOK_000011092	genus	Skeletonema	0.9042	Chromalveolata	Stramenopiles	Bacillariophyta	0	1.521	39.616
SHOK_000016252	order	Phyllodocida	0.8088	Opisthokonta	Metazoa	Annelida	0	1.512	41.311
SHOK_000000218	superkingdom	Eukaryota	0.7688				2.288	0.933	42.877
SHOK_000008956	species	Ophiocoma echinata	0.9904	Opisthokonta	Metazoa	Echinodermata	1.126	0.384	44.187
SHOK_000000813	species	Micromonas commoda	0.8589	Archaeplastida	Viridiplantae	Chlorophyta	0.972	1.637	45.385
SHOK_000001496	superkingdom	Eukaryota	0.8435				0.103	1.147	46.556
SHOK_000000507	class	Mamiellophyceae	0.8062	Archaeplastida	Viridiplantae	Chlorophyta	1.83	1.218	47.723
SHOK_000002291	class	Oomycetes	0.8297	Chromalveolata	Stramenopiles	Oomycota	1.141	0.332	48.713
SHOK_000015585	genus	Skeletonema	0.879	Chromalveolata	Stramenopiles	Bacillariophyta	0.101	0.956	49.673
SHOK_000003880	genus	Ophiocoma	0.8313	Opisthokonta	Metazoa	Echinodermata	0	0.84	50.616
SHOK_000000875	order	Phyllodocida	0.8156	Opisthokonta	Metazoa	Annelida	0	0.787	51.498
SHOK_000000989	superkingdom	Eukaryota	0.7895				0.871	0.622	52.333
SHOK_000000398	phylum	Arthropoda	0.787	Opisthokonta	Metazoa	Arthropoda	0.011	0.712	53.123
SHOK_000001659	class	Dinophyceae	0.8851	Chromalveolata	Alveolata	Dinoflagellata	0.693	0.005	53.896
SHOK_002651932	class	Maxillopoda	0.7746	Opisthokonta	Metazoa	Arthropoda	0.686	0.005	54.668
SHOK_000003678	genus	Aspergillus	0.9233	Opisthokonta	Fungi	Ascomycota	0.004	0.616	55.356
SHOK_000053868	subspecies	Lytechinus variegatus variegatus	0.9968	Opisthokonta	Metazoa	Echinodermata	0.475	0.326	56.028
SHOK_000001136	superkingdom	Eukaryota	0.7613				0.746	0.28	56.67
SHOK_000007936	species	Apionsoma misakianum	0.8286	Opisthokonta	Metazoa	Annelida	0.327	0.383	57.309
SHOK_000011762	order	Terebellida	0.8265	Opisthokonta	Metazoa	Annelida	0.552	0	57.928

Supplementary material 4.3 Results SIMPER analysis for the Discovery Bay vs Montego Bay sites in Jamaica

id	rank	scientific_name	best_id	superkingdom	kingdom	phylum	av_Mangrove	av_Reef	cumsum
SHOK_000000703	superkingdom	Eukaryota	0.7645				6.69	50.478	25.724
SHOK_000000152	class	Dinophyceae	0.9778	Chromalveolata	Alveolata	Dinoflagellata	15.591	3.371	33.093
SHOK_000000486	superkingdom	Eukaryota	0.76				7.423	0	37.454
SHOK_000000184	class	Bivalvia	0.8217	Opisthokonta	Metazoa	Mollusca	7.318	0	41.753
SHOK_000005316	no rank	Eumetazoa	0.7695	Opisthokonta	Metazoa		3.88	0	44.032
SHOK_000001753	no rank	Micromonas pusilla CCMP1545	0.7631	Archaeplastida	Viridiplantae	Chlorophyta	2.852	0.011	45.704
SHOK_000006314	superkingdom	Eukaryota	0.7778				2.063	0	46.916
SHOK_000002150	species	Heterocapsa circularisquama	0.8339	Chromalveolata	Alveolata	Dinoflagellata	1.841	0.005	47.997
SHOK_000000001	superkingdom	Eukaryota	0.7563				0.28	1.811	48.897
SHOK_000001190	superkingdom	Eukaryota	0.753				0.191	1.7	49.784
SHOK_000001659	class	Dinophyceae	0.8851	Chromalveolata	Alveolata	Dinoflagellata	1.479	0.025	50.643
SHOK_000000966	superkingdom	Eukaryota	0.7671				0.244	1.699	51.497
SHOK_000000218	superkingdom	Eukaryota	0.7688				0.941	0.832	52.344
SHOK_000000279	superkingdom	Eukaryota	0.7477				1.402	0	53.167
SHOK_000001282	superkingdom	Eukaryota	0.8025				1.291	0.043	53.908
SHOK_000002814	no rank	Eumetazoa	0.8106	Opisthokonta	Metazoa		0	1.228	54.629
SHOK_000013576	kingdom	Metazoa	0.7554	Opisthokonta	Metazoa		0.142	1.299	55.309
SHOK_000000848	superkingdom	Eukaryota	0.7576				0.2	1.354	55.986
SHOK_000000506	order	Terebellida	0.8182	Opisthokonta	Metazoa	Annelida	1.144	0	56.659
SHOK_000029624	suborder	Aplousobranchia	0.7962	Opisthokonta	Metazoa	Chordata	1.135	0.263	57.317
SHOK_000001742	superkingdom	Eukaryota	0.7695				0.161	1.237	57.949
SHOK_000009258	superkingdom	Eukaryota	0.7822				0	1.054	58.569
SHOK_000005148	genus	Heterocapsa	0.8179	Chromalveolata	Alveolata	Dinoflagellata	1.066	0.043	59.181
SHOK_000000932	superkingdom	Eukaryota	0.8261				1.019	0	59.78
SHOK_000028083	kingdom	Metazoa	0.7771	Opisthokonta	Metazoa		0.916	0.03	60.314
SHOK_000000197	superkingdom	Eukaryota	0.7938				0	0.895	60.84
SHOK_000000898	kingdom	Stramenopiles	0.8025	Chromalveolata	Stramenopiles		1.094	0.366	61.362
SHOK_000000230	no rank	Eumetazoa	0.8333	Opisthokonta	Metazoa		0	0.865	61.871
SHOK_000000523	kingdom	Stramenopiles	0.8287	Chromalveolata	Stramenopiles		1.036	0.601	62.376
SHOK_000001012	species	Cunina fowleri	0.7981	Opisthokonta	Metazoa	Cnidaria	0	0.861	62.882

Supplementary material 4.3 Results SIMPER analysis for the mangrove vs reef sites in Turks & Caicos

Class	Family	Species	Class	Family	Species
Actinopterygii	Acipenseridae	*Acipenser sturio*	Actinopterygii	Gobiidae	*Pomatoschistus microps*
Actinopterygii	Agonidae	*Agonus cataphractus*	Actinopterygii	Gobiidae	*Pomatoschistus minutus*
Actinopterygii	Ammodytidae	*Ammodytes marinus*	Actinopterygii	Gobiidae	*Pomatoschistus norvegicus*
Actinopterygii	Ammodytidae	*Ammodytes tobianus*	Actinopterygii	Gobiidae	*Pomatoschistus pictus*
Actinopterygii	Ammodytidae	*Hyperoplus immaculatus*	Actinopterygii	Gobiidae	*Thorogobius ephippiatus*
Actinopterygii	Ammodytidae	*Hyperoplus lanceolatus*	Actinopterygii	Labridae	*Centrolabrus exoletus*
Actinopterygii	Anarhichadidae	*Anarhichas lupus*	Actinopterygii	Labridae	*Ctenolabrus rupestris*
Actinopterygii	Anguillidae	*Anguilla anguilla*	Actinopterygii	Labridae	*Labrus bergylta*
Actinopterygii	Argentinidae	*Argentina sphyraena*	Actinopterygii	Labridae	*Labrus mixtus*
Actinopterygii	Atherinidae	*Atherina boyeri*	Actinopterygii	Labridae	*Symphodus melops*
Actinopterygii	Atherinidae	*Atherina presbyter*	Actinopterygii	Liparidae	*Liparis liparis*
Actinopterygii	Balistidae	*Balistes capriscus*	Actinopterygii	Liparidae	*Liparis montagui*
Actinopterygii	Belonidae	*Belone belone*	Actinopterygii	Lophiidae	*Lophius piscatorius*
Actinopterygii	Belonidae	*Scomberesox saurus*	Actinopterygii	Lotidae	*Ciliata mustela*
Actinopterygii	Blenniidae	*Blennius ocellaris*	Actinopterygii	Lotidae	*Ciliata septentrionalis*
Actinopterygii	Blenniidae	*Coryphoblennius galerita*	Actinopterygii	Lotidae	*Molva molva*
Actinopterygii	Blenniidae	*Lipophrys pholis*	Actinopterygii	Merlucciidae	*Merluccius merluccius*
Actinopterygii	Blenniidae	*Parablennius gattorugine*	Actinopterygii	Moronidae	*Dicentrarchus labrax*
Actinopterygii	Bothidae	*Arnoglossus laterna*	Actinopterygii	Mugilidae	*Chelon labrosus*
Actinopterygii	Callionymidae	*Callionymus lyra*	Actinopterygii	Mugilidae	*Liza aurata*
Actinopterygii	Callionymidae	*Callionymus maculatus*	Actinopterygii	Mugilidae	*Liza ramada*
Actinopterygii	Callionymidae	*Callionymus reticulatus*	Actinopterygii	Mullidae	*Mullus surmuletus*
Actinopterygii	Carangidae	*Trachurus trachurus*	Actinopterygii	Nemacheilidae	*Barbatula barbatula*
Actinopterygii	Cepolidae	*Cepola macrophthalma*	Actinopterygii	Osmeridae	*Osmerus eperlanus*
Actinopterygii	Clupeidae	*Alosa alosa*	Actinopterygii	Percidae	*Gymnocephalus cernuus*
Actinopterygii	Clupeidae	*Alosa fallax*	Actinopterygii	Percidae	*Perca fluviatilis*
Actinopterygii	Clupeidae	*Clupea harengus*	Actinopterygii	Percidae	*Sander lucioperca*
Actinopterygii	Clupeidae	*Sardina pilchardus*	Actinopterygii	Pholidae	*Pholis gunnellus*
Actinopterygii	Clupeidae	*Sprattus sprattus*	Actinopterygii	Phycidae	*Phycis blennoides*
Actinopterygii	Congridae	*Conger conger*	Actinopterygii	Pleuronectidae	*Glyptocephalus cynoglossus*
Actinopterygii	Cottidae	*Cottus gobio*	Actinopterygii	Pleuronectidae	*Hippoglossoides platessoides*
Actinopterygii	Cottidae	*Micrenophrys lilljeborgii*	Actinopterygii	Pleuronectidae	*Hippoglossus hippoglossus*
Actinopterygii	Cottidae	*Myoxocephalus scorpius*	Actinopterygii	Pleuronectidae	*Limanda limanda*
Actinopterygii	Cottidae	*Taurulus bubalis*	Actinopterygii	Pleuronectidae	*Microstomus kitt*
Actinopterygii	Cyclopteridae	*Cyclopterus lumpus*	Actinopterygii	Pleuronectidae	*Platichthys flesus*
Actinopterygii	Cyprinidae	*Abramis brama*	Actinopterygii	Pleuronectidae	*Pleuronectes platessa*
Actinopterygii	Cyprinidae	*Alburnus alburnus*	Actinopterygii	Salmonidae	*Coregonus albula*
Actinopterygii	Cyprinidae	*Barbus barbus*	Actinopterygii	Salmonidae	*Coregonus autumnalis*
Actinopterygii	Cyprinidae	*Blicca bjoerkna*	Actinopterygii	Salmonidae	*Coregonus lavaretus*
Actinopterygii	Cyprinidae	*Carassius auratus*	Actinopterygii	Salmonidae	*Coregonus oxyrinchus*
Actinopterygii	Cyprinidae	*Cyprinus carpio*	Actinopterygii	Salmonidae	*Oncorhynchus mykiss*
Actinopterygii	Cyprinidae	*Gobio gobio*	Actinopterygii	Salmonidae	*Salmo salar*
Actinopterygii	Cyprinidae	*Leuciscus cephalus*	Actinopterygii	Salmonidae	*Salmo trutta*
Actinopterygii	Cyprinidae	*Leuciscus leuciscus*	Actinopterygii	Salmonidae	*Salvelinus alpinus*
Actinopterygii	Cyprinidae	*Phoxinus phoxinus*	Actinopterygii	Salmonidae	*Thymallus thymallus*
Actinopterygii	Cyprinidae	*Rutilus rutilus*	Actinopterygii	Sciaenidae	*Argyrosomus regius*
Actinopterygii	Cyprinidae	*Scardinius erythrophthalmus*	Actinopterygii	Scombridae	*Scomber scombrus*
Actinopterygii	Cyprinidae	*Tinca tinca*	Actinopterygii	Scophthalmidae	*Lepidorhombus whiffiagonis*
Actinopterygii	Engraulidae	*Engraulis encrasicolus*	Actinopterygii	Scophthalmidae	*Phrynorhombus norvegicus*
Actinopterygii	Esocidae	*Esox lucius*	Actinopterygii	Scophthalmidae	*Scophthalmus maximus*
Actinopterygii	Gadidae	*Gadiculus argenteus*	Actinopterygii	Scophthalmidae	*Scophthalmus rhombus*
Actinopterygii	Gadidae	*Gadus morhua*	Actinopterygii	Scophthalmidae	*Zeugopterus punctatus*
Actinopterygii	Gadidae	*Melanogrammus aeglefinus*	Actinopterygii	Scophthalmidae	*Zeugopterus regius*
Actinopterygii	Gadidae	*Merlangius merlangus*	Actinopterygii	Scorpaenidae	*Scorpaena porcus*
Actinopterygii	Gadidae	*Micromesistius poutassou*	Actinopterygii	Soleidae	*Buglossidium luteum*
Actinopterygii	Gadidae	*Pollachius pollachius*	Actinopterygii	Soleidae	*Microchirus variegatus*
Actinopterygii	Gadidae	*Pollachius virens*	Actinopterygii	Soleidae	*Pegusa lascaris*
Actinopterygii	Gadidae	*Raniceps raninus*	Actinopterygii	Soleidae	*Solea solea*
Actinopterygii	Gadidae	*Trisopterus esmarkii*	Actinopterygii	Sparidae	*Pagellus bogaraveo*
Actinopterygii	Gadidae	*Trisopterus luscus*	Actinopterygii	Sparidae	*Sparus aurata*
Actinopterygii	Gadidae	*Trisopterus minutus*	Actinopterygii	Sparidae	*Spondyliosoma cantharus*
Actinopterygii	Gaidropsaridae	*Enchelyopus cimbrius*	Actinopterygii	Sternoptychidae	*Maurolicus muelleri*
Actinopterygii	Gaidropsaridae	*Gaidropsarus mediterraneus*	Actinopterygii	Stichaeidae	*Chirolophis ascanii*
Actinopterygii	Gaidropsaridae	*Gaidropsarus vulgaris*	Actinopterygii	Stichaeidae	*Lumpenus lampretaeformis*
Actinopterygii	Gasterosteidae	*Gasterosteus aculeatus*	Actinopterygii	Syngnathidae	*Entelurus aequoreus*
Actinopterygii	Gasterosteidae	*Pungitius pungitius*	Actinopterygii	Syngnathidae	*Hippocampus guttulatus*
Actinopterygii	Gasterosteidae	*Spinachia spinachia*	Actinopterygii	Syngnathidae	*Hippocampus hippocampus*
Actinopterygii	Gobiesocidae	*Apletodon dentatus*	Actinopterygii	Syngnathidae	*Nerophis lumbriciformis*
Actinopterygii	Gobiesocidae	*Diplecogaster bimaculata*	Actinopterygii	Syngnathidae	*Nerophis ophidion*
Actinopterygii	Gobiesocidae	*Lepadogaster lepadogaster*	Actinopterygii	Syngnathidae	*Syngnathus acus*
Actinopterygii	Gobiidae	*Aphia minuta*	Actinopterygii	Syngnathidae	*Syngnathus rostellatus*
Actinopterygii	Gobiidae	*Buenia jeffreysii*	Actinopterygii	Syngnathidae	*Syngnathus typhle*
Actinopterygii	Gobiidae	*Crystallogobius linearis*	Actinopterygii	Trachinidae	*Echiichthys vipera*
Actinopterygii	Gobiidae	*Gobius cobitis*	Actinopterygii	Trachinidae	*Trachinus draco*
Actinopterygii	Gobiidae	*Gobius niger*	Actinopterygii	Triglidae	*Chelidonichthys cuculus*
Actinopterygii	Gobiidae	*Gobius paganellus*	Actinopterygii	Triglidae	*Chelidonichthys lucerna*
Actinopterygii	Gobiidae	*Gobiusculus flavescens*	Actinopterygii	Triglidae	*Eutrigla gurnardus*
Actinopterygii	Gobiidae	*Lesueurigobius friesii*	Actinopterygii	Triglidae	*Trigla lyra*
Actinopterygii	Gobiidae	*Pomatoschistus lozanoi*	Actinopterygii	Triglidae	*Trigloporus lastoviza*
			Actinopterygii	Zeidae	*Zeus faber*
			Actinopterygii	Zoarcidae	*Zoarces viviparus*

Supplementary material 5.1 List of 160 teleost species that are found in UK transitional coastal waters, provided by the Water Framework Directive, United Kingdom Technical Advisory Group

References

Alberdi, A. et al., 2017. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, 2017(May), pp.1–14.

Amberg, J.J. et al., 2015. Improving efficiency and reliability of environmental DNA analysis for silver carp. *Journal of Great Lakes Research*, 41(2), pp.367–373. Available at: http://www.sciencedirect.com/science/article/pii/S038013301500057X.

Andruszkiewicz, E.A. et al., 2017. Persistence of marine fish environmental DNA and the influence of sunlight H. Doi, ed. *PLOS ONE*, 12(9), p.e0185043. Available at: http://dx.plos.org/10.1371/journal.pone.0185043 [Accessed September 16, 2017].

Anon, FurlanGleeson2016MFR.

Appeltans, W. et al., 2012. The magnitude of global marine species diversity. *Current Biology*, 22(23), pp.2189–2202.

Aylagas, E. et al., 2016. Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science*, 3(June), pp.1–12. Available at: http://journal.frontiersin.org/Article/10.3389/fmars.2016.00096/abstract.

Bailiff, M.D. & Karl, D.M., 1991. Dissolved and particulate DNA dynamics during a spring bloom in the Antarctic Peninsula region, 1986–1987. *Deep Sea Research Part A. Oceanographic Research Papers*, 38(8–9), pp.1077–1095. Available at: https://www.sciencedirect.com/science/article/pii/019801499190097Y [Accessed December 18, 2017].

Baird, D.J., Hajibabaei, M. & Brunswick, N., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, 21(8), pp.2039–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22590728.

Bakker, J. et al., 2017. Environmental DNA reveals tropical shark diversity and abundance in contrasting levels of anthropogenic impact. *Scientific Reports*, (May), pp.1–11. Available at: http://dx.doi.org/10.1038/s41598-017-17150-2.

De Barba, M. et al., 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources*, 14(2), pp.306–323. Available at: http://dx.doi.org/10.1111/1755-0998.12188.

Barnes, M.A. et al., 2014a. Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environmental Science & Technology*, 48(3), pp.1819–1827. Available at: http://pubs.acs.org/doi/abs/10.1021/es404734p.

Barnes, M.A. et al., 2014b. Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environmental Science & Technology*, 48(3), pp.1819–1827. Available at: http://pubs.acs.org/doi/abs/10.1021/es404734p.

Barnes, M.A. & Turner, C.R., 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17(1), pp.1–17.

Bauer, H. et al., 2015. Lion (*Panthera leo*) populations are declining rapidly across Africa, except in intensively managed areas. *Proceedings of the National Academy of Sciences*, 112(48), pp.14894–14899. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1500664112.

Baum, J.K. & Worm, B., 2009. Cascading top-down effects of changing oceanic predator abundances. *Journal of Animal Ecology*, 78(4), pp.699–714.

Bensasson, D. et al., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in ecology & evolution*, 16(6), pp.314–321.

Bik, H.M. et al., 2012. Dramatic shifts in benthic microbial eukaryote communities following the deepwater horizon oil spill. *PLoS ONE*, 7(6).

Boettiger, C., Lang, D.T. & Wainwright, P.C., 2012. rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology*, 81(6), pp.2030–2039. Available at: http://dx.doi.org/10.1111/j.1095-8649.2012.03464.x.

Bohmann, K. et al., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, 29(6), pp.358–367. Available at: http://dx.doi.org/10.1016/j.tree.2014.04.003.

Bond, M.E. et al., 2012. Reef Sharks Exhibit Site-Fidelity and Higher Relative Abundance in Marine Reserves on the Mesoamerican Barrier Reef V. Laudet, ed.

PLoS ONE, 7(3), p.e32983. Available at: http://dx.plos.org/10.1371/journal.pone.0032983.

Bonfil, R., 1994. *Overview of world elasmobranch fisheries. FAO. Fisheries Technical Paper 341*, Rome. Available at: http://www.fao.org/3/a-v3210e/index.html%5Cnhttp://www.fao.org/documents/card/en/c/a1f94b92-d9a6-5971-9062-34dd9eb153f7/.

BOUCHET, P. et al., 2002. Assessing the magnitude of species richness in tropical marine environments: exceptionally high numbers of molluscs at a New Caledonia site. *Biological Journal of the Linnean Society*, 75(4), pp.421–436. Available at: http://dx.doi.org/10.1046/j.1095-8312.2002.00052.x.

Bourlat, S.J. et al., 2013. Genomics in marine monitoring: New opportunities for assessing marine health status. *Marine Pollution Bulletin*, 74(1), pp.19–31. Available at: http://dx.doi.org/10.1016/j.marpolbul.2013.05.042.

Boyer, F. et al., 2016. obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16(1), pp.176–182.

Boyer, S. et al., 2012. Sliding window analyses for optimal selection of mini-barcodes, and application to 454-pyrosequencing for specimen identification from degraded DNA. *PLoS ONE*, 7(5).

Brandl, S. et al., 2015. Ten real-time PCR assays for detection of fish predation at the community level in the San Francisco Estuary–Delta. *Molecular Ecology Resources*, 15(2), pp.278–284.

Breier, J.A. et al., 2014. A large volume particulate and water multi-sampler with in situ preservation for microbial and biogeochemical studies. *Deep-Sea Research Part I: Oceanographic Research Papers*, 94, pp.195–206.

Britton, J.R., Davies, G.D. & Harrod, C., 2010. Trophic interactions and consequent impacts of the invasive fish Pseudorasbora parva in a native aquatic foodweb: A field investigation in the UK. *Biological Invasions*, 12(6), pp.1533–1542.

BROWN, S.D.J. et al., 2012. Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, 12(3), pp.562–565. Available at: http://dx.doi.org/10.1111/j.1755-0998.2011.03108.x.

Bucklin, A. (Department of M.S.-U. of C. et al., 2016. Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of Plankton Research*, 0(0), pp.1–8. Available at: http://plankt.oxfordjournals.org/content/early/2016/04/06/plankt.fbw023.abstract.

Bucklin, A., Steinke, D. & Blanco-Bercial, L., 2011. DNA barcoding of marine metazoa. *Annual review of marine science*, 3, pp.471–508.

Burgar, J.M. et al., 2014. Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. *Molecular Ecology*, 23(15), pp.3605–3617. Available at: http://dx.doi.org/10.1111/mec.12531.

Buxton, A.S. et al., 2017. Seasonal variation in environmental DNA in relation to population size and environmental factors. *Scientific Reports*, 7(1), p.46294. Available at: http://www.nature.com/articles/srep46294 [Accessed July 30, 2017].

Bylemans, J. et al., 2017. An environmental DNA-based method for monitoring spawning activity: a case study, using the endangered Macquarie perch (Macquaria australasica). *Methods in Ecology and Evolution*, 8(5), pp.646–655.

Camhi, M.D. et al., 2009. The conservation status of pelagic sharks and rays, IUCN Species Survival Commission's Shark Specialist Group. , pp.19–23. Available at: http://stage.pewenvironment.com/uploadedFiles/PEG/Newsroom/Press_Release/The_Conservation_Status_of_Pelagic_Sharks_and_Rays_Hand_out_Report_of_the_IUCN_Shark_Specialist_Group_Pelagic_Shark_Red_List_Workshop.pdf.

Cave, L. et al., 2016. Efficiency and sensitivity of the digital droplet PCR for the quantification of antibiotic resistance genes in soils and organic residues. *Applied microbiology and biotechnology*, 100(24), pp.10597–10608.

Chain, F.J.J. et al., 2016. Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports R. Cowie, ed. *Diversity and Distributions*, 22(5), pp.493–504. Available at: http://doi.wiley.com/10.1111/ddi.12427 [Accessed February 9, 2016].

Chapman, D.D. et al., 2013. Give Shark Sanctuaries a Chance. *Science (New York, N.Y.)*, 339(6121), pp.757–757. Available at:

http://science.sciencemag.org/content/339/6121/757.1.abstract.

Cinner, J.E. et al., 2012. Comanagement of coral reef social-ecological systems. *Proceedings of the National Academy of Sciences*, 109(14), pp.5219–5222.

Cinner, J.E. et al., 2013. Evaluating Social and Ecological Vulnerability of Coral Reef Fisheries to Climate Change S. Dupont, ed. *PLoS ONE*, 8(9), p.e74321. Available at: http://dx.plos.org/10.1371/journal.pone.0074321.

Cinner, J.E., McClanahan & T R, 2006. Socioeconomic factors that lead to overfishing in small-scale coral reef fisheries of Papua New Guinea. *Environmental Conservation*, 33(1), pp.73–80.

Civade, R. et al., 2016. Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System C. Garcia de Leaniz, ed. *PLOS ONE*, 11(6), p.e0157366. Available at: http://dx.plos.org/10.1371/journal.pone.0157366.

Clarke, L.J. et al., 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, 7(3), pp.873–883.

Clarke, L.J. et al., 2014. Environmental metabarcodes for insects: in silicoPCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, 14(6), pp.1160–1170. Available at: http://dx.doi.org/10.1111/1755-0998.12265.

Coissac, E., Riaz, T. & Puillandre, N., 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, 21(8), pp.1834–1847.

Compagno, L.J. V, Dando, M. & Fowler, S.L., 2005. *A Field Guide to the Sharks of the World*, Collins. Available at: https://books.google.co.uk/books?id=k_-_XEyQmsQC.

Comtet, T. et al., 2015. DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, 17(3), pp.905–922. Available at: http://link.springer.com/10.1007/s10530-015-0854-y [Accessed July 11, 2016].

Cordier, T. et al., 2017. Predicting the Ecological Quality Status of Marine Environments from eDNA Metabarcoding Data Using Supervised Machine Learning. *Environmental Science & Technology*, 51(16), pp.9118–9126.

Available at: http://dx.doi.org/10.1021/acs.est.7b01518.

Costello, C. et al., 2012. Status and Solutions for the World's Unassessed Fisheries. *Science*, 338(6106), p.517 LP-520. Available at: http://science.sciencemag.org/content/338/6106/517.abstract.

D'agata, S. et al., 2016. Marine reserves lag behind wilderness in the conservation of key functional roles. *Nature Communications*, 7(May), p.12000. Available at: http://dx.doi.org/10.1038/ncomms12000%5Cnhttp://10.1038/ncomms12000.

Darling, J. a. & Mahon, A.R., 2011. From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research*, 111(7), pp.978–988. Available at: http://dx.doi.org/10.1016/j.envres.2011.02.001.

Davidson, L.N.K. & Dulvy, N.K., 2017. Global marine protected areas for avoiding extinctions. *Nature Ecology and Evolution*, 1(January), pp.1–6. Available at: http://dx.doi.org/10.1038/s41559-016-0040.

Davison, P.I. et al., 2017. Application of environmental DNA analysis to inform invasive fish eradication operations. *The Science of Nature*, 104(3–4), p.35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28357478 [Accessed April 28, 2017].

Davy, C.M., Kidd, A.G. & Wilson, C.C., 2015. Development and Validation of Environmental DNA (eDNA) Markers for Detection of Freshwater Turtles A. R. Mahon, ed. *PLOS ONE*, 10(7), p.e0130965. Available at: http://dx.plos.org/10.1371/journal.pone.0130965 [Accessed June 29, 2016].

Deagle, B.E. et al., 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology letters*, 10(9), pp.1789–1793. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25209199%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4190964.

Deagle, B.E. et al., 2017. Genetic monitoring of open ocean biodiversity: an evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Molecular Ecology Resources*, 12(10), pp.3218–3221. Available at: http://doi.wiley.com/10.1111/1755-0998.12740.

Deiner, K. et al., 2015. Choice of capture and extraction methods affect detection of

freshwater biodiversity from environmental DNA. *Biological Conservation*, 183, pp.53–63. Available at: http://dx.doi.org/10.1016/j.biocon.2014.11.018.

Deiner, K. et al., 2016. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications*, 7(0), p.12544. Available at: http://biorxiv.org/content/early/2015/06/11/020800.abstract.

Deiner, K. et al., 2017. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*, 2017(April), pp.1–11.

Deiner, K. & Altermatt, F., 2014. Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, 9(2).

Dejean, T. et al., 2012. Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog Lithobates catesbeianus. *Journal of Applied Ecology*, 49(4), pp.953–959. Available at: http://doi.wiley.com/10.1111/j.1365-2664.2012.02171.x.

Dejean, T. et al., 2011. Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE*, 6(8), pp.8–11.

Dell'Anno, A. & Corinaldesi, C., 2004. Degradation and Turnover of Extracellular DNA in Marine Sediments : Ecological and Methodological Considerations Degradation and Turnover of Extracellular DNA in Marine Sediments : Ecological and Methodological Considerations. *Applied and environmental microbiology*, 70(7), pp.4384–4386.

Dirzo, R. et al., 2014. Defaunation in the Anthropocene. *Science*, 345(6195), pp.401–406.

Djurhuus, A. et al., 2017. Evaluation of Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels. *Frontiers in Marine Science*, 4, p.314. Available at: http://journal.frontiersin.org/article/10.3389/fmars.2017.00314/full [Accessed November 20, 2017].

Djurhuus, A. et al., 2018. Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. *Limnology and Oceanography: Methods*, (Schminke 2007). Available at:

http://doi.wiley.com/10.1002/lom3.10237.

Doi, H., Takahara, T., et al., 2015. Droplet digital polymerase chain reaction (PCR) outperforms real-time PCR in the detection of environmental DNA from an invasive fish species. *Environmental science & technology*, 49(9), pp.5601–8. Available at: http://pubs.acs.org.cyber.usask.ca/doi/abs/10.1021/acs.est.5b00253.

Doi, H. et al., 2017. Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology*, 62(1), pp.30–39.

Doi, H., Uchii, K., et al., 2015. Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLoS ONE*, 10(3), pp.1–11.

Doyle, J.R., McKinnon, A.D. & Uthicke, S., 2017. Quantifying larvae of the coralivorous seastar Acanthaster cf. solaris on the Great Barrier Reef using qPCR. *Marine Biology*, 164(8), p.176. Available at: https://doi.org/10.1007/s00227-017-3206-x.

Drummond, A.J. et al., 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, 4(1), p.46. Available at: http://www.gigasciencejournal.com/content/4/1/46.

Dulvy, N.K. et al., 2014. Extinction risk and conservation of the world's sharks and rays. *eLife*, 3, p.e00590.

Dulvy, N.K. et al., 2016. Ghosts of the coast: Global extinction risk and conservation of sawfishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 26(1), pp.134–153.

Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics (Oxford, England)*, 14(9), pp.755–763.

Edgar, G.J. et al., 2014. Global conservation outcomes depend on marine protected areas with five key features. *Nature*, 506(7487), pp.216–220.

Edgar, R.C. et al., 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16), pp.2194–2200.

Eichmiller, J.J., Bajer, P.G. & Sorensen, P.W., 2014. The Relationship between the Distribution of Common Carp and Their Environmental DNA in a Small Lake. *PLoS ONE*, 9(11), p.e112611. Available at:

http://dx.plos.org/10.1371/journal.pone.0112611.

Eichmiller, J.J., Miller, L.M. & Sorensen, P.W., 2016. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources*, 16(1), pp.56–68.

Elbrecht, V. & Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, 10(7), pp.1–16.

Elbrecht, V. & Leese, F., 2016. Development and validation of DNA metabarcoding COI primers for aquatic invertebrates using the R package " PrimerMiner ." *unknown - in peer review. Kerry walsh sent through*, (May).

Estes, J.A. et al., 2011. Trophic Downgrading of Planet Earth. *Science (New York, N.Y.)*, 333(6040), pp.301–306.

Evans, N.T. et al., 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, 16(1), pp.29–41.

Evans, N.T. & Lamberti, G.A., 2017. Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool. *Fisheries Research*, (February), pp.0–1. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0165783617302618.

Ferretti, F. et al., 2010. Patterns and ecosystem consequences of shark declines in the ocean. *Ecology Letters*, 13(8), pp.1055–1071.

Ficetola, G.F. et al., 2010. An In silico approach for the evaluation of DNA barcodes. *BMC Genomics*, 11(1), p.434. Available at: http://dx.doi.org/10.1186/1471-2164-11-434.

Ficetola, G.F. et al., 2008. Species detection using environmental DNA from water samples. *Biol Lett*, 4. Available at: http://dx.doi.org/10.1098/rsbl.2008.0118.

Ficetola, G.F.G.F. et al., 2008. Species detection using environmental DNA from water samples. *Biology Letters*, 4(4), pp.423–425. Available at: http://rsbl.royalsocietypublishing.org/cgi/content/abstract/4/4/423%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2610135&tool=pmcentrez&rendertype=abstract.

Fields, A.T. et al., 2015. A Novel Mini-DNA Barcoding Assay to Identify Processed Fins from Internationally Protected Shark Species A. Stow, ed. *PLOS ONE*, 10(2), p.e0114844. Available at: http://dx.plos.org/10.1371/journal.pone.0114844.

Foissner, W., 2008. Protist diversity and distribution: Some basic considerations. *Biodiversity and Conservation*, 17(2), pp.235–242.

Folmer, O. et al., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), pp.294–299.

Fonseca, V.G. et al., 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature communications*, 1(7), p.98. Available at: http://dx.doi.org/10.1038/ncomms1095.

Foote, A.D. et al., 2012. Investigating the Potential Use of Environmental DNA (eDNA) for Genetic Monitoring of Marine Mammals. *PLoS ONE*, 7(8), pp.2–7.

Fraga, S. et al., 1995. Gyrodinium impudicum sp. nov. (Dinophyceae), a non toxic, chain-forming, red tide dinoflagellate. *Phycologia*, 34(6), pp.514–521. Available at: https://doi.org/10.2216/i0031-8884-34-6-514.1.

French, S.S. et al., 2011. Human disturbance influences reproductive success and growth rate in California sea lions (Zalophus californianus). *PLoS ONE*, 6(3).

Frid, A. & Dill, L., 2002. Human-caused disturbance stimuli as a form of predation risk. *Ecology and Society*, 6(1).

Fu, X.H. et al., 2012. Persistence and renaturation efficiency of thermally treated waste recombinant DNA in defined aquatic microcosms. *Journal of environmental science and health. Part A, Toxic/hazardous substances & environmental engineering*, 47(13), pp.1975–1983.

Fukumoto, S., Ushimaru, A. & Minamoto, T., 2015. A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: a case study of giant salamanders in Japan E. Crispo, ed. *Journal of Applied Ecology*, 52(2), pp.358–365. Available at: http://doi.wiley.com/10.1111/1365-2664.12392 [Accessed July 8, 2016].

Furlan, E.M. et al., 2016. A framework for estimating the sensitivity of eDNA

surveys. *Molecular Ecology Resources*, 16(3), pp.641–654.

Garcia, V.B., Lucifora, L.O. & Myers, R.A., 2008. The importance of habitat and life history to extinction risk in sharks, skates, rays and chimaeras. *Proc Biol Sci*, 275(1630), pp.83–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17956843.

Gargan, L.M. et al., 2017. Development of a sensitive detection method to survey pelagic biodiversity using eDNA and quantitative PCR: a case study of devil ray at seamounts. *Marine Biology*, 164(5), p.112. Available at: http://link.springer.com/10.1007/s00227-017-3141-x [Accessed April 24, 2017].

Geller, J. et al., 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, 13(5), pp.851–861.

Giles, R.E. et al., 1980. Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 77(11), pp.6715–6719. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=350359&tool=pmcentrez&rendertype=abstract.

Gilman, E. et al., 2008. Shark interactions in pelagic longline fisheries. *Marine Policy*, 32(1), pp.1–18.

Goetze, J.S. et al., 2017. Fish wariness is a more sensitive indicator to changes in fishing pressure than abundance, length or biomass. *Ecological Applications*, 27(4), pp.1178–1189. Available at: http://dx.doi.org/10.1002/eap.1511.

Goldberg, C.S. et al., 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, pp.1299–1307. Available at: http://doi.wiley.com/10.1111/2041-210X.12595.

Goldberg, C.S. et al., 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32(3), pp.792–800. Available at: http://www.journals.uchicago.edu/doi/10.1899/13-046.1.

Goldberg, C.S. et al., 2011. Molecular detection of vertebrates in stream water: A

demonstration using rocky mountain tailed frogs and Idaho giant salamanders. *PLoS ONE*, 6(7), p.e22746. Available at: http://dx.plos.org/10.1371/journal.pone.0022746.

Goodwin, K.D. et al., 2017. DNA Sequencing as a Tool to Monitor Marine Ecological Status. *Frontiers in Marine Science*, 4(May), pp.1–14. Available at: http://journal.frontiersin.org/article/10.3389/fmars.2017.00107/full.

Gotelli, N.J. & Colwell, R.K., 2001. Quantifyinf Biodiversity: Procedures and Pitfalls in the Measurement and Comparison of Species Richness. *Ecology Letters*, 4(4), pp.379–391.

Graham, N. a J. & Mcclanahan, T.R., 2013. The Last Call for Marine Wilderness? *BioScience*, 63(5), pp.397–402. Available at: http://bioscience.oxfordjournals.org/cgi/doi/10.1525/bio.2013.63.5.13.

Guardiola, M. et al., 2015. Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons S. Duperron, ed. *PLOS ONE*, 10(10), p.e0139633. Available at: http://dx.plos.org/10.1371/journal.pone.0139633.

Guil, N., 2011. Molecular approach to micrometazoans. Are they here, there and everywhere? In D. Fontaneto, ed. *Biogeography of Microscopic Organisms: Is Everything Small Everywhere?*. Systematics Association Special Volume Series. Cambridge: Cambridge University Press, pp. 284–306. Available at: https://www.cambridge.org/core/books/biogeography-of-microscopic-organisms/molecular-approach-to-micrometazoans-are-they-here-there-and-everywhere/5D62F14BC52A3969D7C59A8C3CA35958.

Gustavson, M.S. et al., 2015. An eDNA assay for Irish Petromyzon marinus and Salmo trutta and field validation in running water. *Journal of Fish Biology*, 87(5), pp.1254–1262. Available at: http://doi.wiley.com/10.1111/jfb.12781 [Accessed October 10, 2016].

Guttridge, T.L. et al., 2017. Philopatry and Regional Connectivity of the Great Hammerhead Shark, Sphyrna mokarran in the U.S. and Bahamas. *Frontiers in Marine Science*, 4, p.3. Available at: http://journal.frontiersin.org/article/10.3389/fmars.2017.00003.

Hajibabaei, M. et al., 2006. A minimalist barcode can identify a specimen whose

DNA is degraded. *Mol Ecol Notes*, 6. Available at: http://dx.doi.org/10.1111/j.1471-8286.2006.01470.x.

Hajibabaei, M. et al., 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology*, 5(1), p.24. Available at: http://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-5-24.

Hajibabaei, M. et al., 2011. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, 6(4).

Hallegraeff, G.M. et al., 2003. *Manual on Harmful Marine Microalgae*, Unesco. Available at: https://books.google.co.uk/books?id=vUsVAQAAIAAJ.

Handelsman, J., 2005. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 69(1), pp.195–195. Available at: http://mmbr.asm.org/cgi/doi/10.1128/MMBR.69.1.195.2005.

Hänfling, B. et al., 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, 25(13), pp.3101–3119. Available at: http://doi.wiley.com/10.1111/mec.13660.

Hansell, A.C. et al., 2017. Local indicators of abundance and demographics for the coastal shark assemblage of Bimini, Bahamas. *Fisheries Research*, (September), pp.1–11. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0165783617302643.

Hardy, C.M. et al., 2011. DNA barcoding to support conservation: species identification, genetic structure and biogeography of fishes in the Murray—Darling River Basin, Australia. *Marine and Freshwater Research*, 62(8), pp.887–901. Available at: https://doi.org/10.1071/MF11027.

Harvey, E.S. et al., 2007. Bait attraction affects the performance of remote underwater video stations in assessment of demersal fish community structure. *Marine Ecology Progress Series*, 350(May 2014), pp.245–254.

Harvey, J.B.J. et al., 2012. Robotic sampling, in situ monitoring and molecular detection of marine zooplankton. *Journal of Experimental Marine Biology and Ecology*, 413, pp.60–70. Available at: http://dx.doi.org/10.1016/j.jembe.2011.11.022.

Hawkins, J.P. & Roberts, C.M., 2004. Effects of artisanal fishing on Caribbean coral reefs. *Conservation Biology*, 18(1), pp.215–226.

Hebert, P.D.N., Cywinska, A., et al., 2003. Biological identifications through DNA barcodes. *Proc R Soc B*, 270. Available at: http://dx.doi.org/10.1098/rspb.2002.2218.

Hebert, P.D.N., Ratnasingham, S. & deWaard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc B*, 270. Available at: http://dx.doi.org/10.1098/rsbl.2003.0025.

Heithaus, M.R., Wirsing, A.J. & Dill, L.M., 2012. The ecological importance of intact top-predator populations: a synthesis of 15 years of research in a seagrass ecosystem. *Marine and Freshwater Research*, 63(11), pp.1039–1050. Available at: https://doi.org/10.1071/MF12024.

Herder, J.E. et al., 2014. Environmental DNA - a review of the possible applications for the\rdetection of (invasive) species. , (October), p.114.

Heupel, M.R. et al., 2014. Sizing up the ecological role of sharks as predators. *Marine Ecology Progress Series*, 495, pp.291–298.

Hindson, B.J. et al., 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, 83(22), pp.8604–8610.

Hinlo, R. et al., 2017. Methods to maximise recovery of environmental DNA from water samples H. Doi, ed. *PLoS ONE*, 12(6), p.e0179251. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28604830 [Accessed June 13, 2017].

Hirai, J. et al., 2015. A metagenetic approach for revealing community structure of marine planktonic copepods. *Molecular ecology resources*, 15(1), pp.68–80.

Hobday, A.J. et al., 2011. Ecological risk assessment for the effects of fishing. *Fisheries Research*, 108(2), pp.372–384. Available at: http://www.sciencedirect.com/science/article/pii/S0165783611000324.

Hoffmann, C., Schubert, G. & Calvignac-Spencer, S., 2016. Aquatic biodiversity assessment for the lazy. *Molecular Ecology*, 25(4), pp.846–848. Available at: http://doi.wiley.com/10.1111/mec.13535 [Accessed February 15, 2016].

Horiguchi, T., 1995. Heterocapsa circularisquama sp. nov. (Peridiniales, Dinophyceae): A new marine dinoflagellate causing mass mortality of bivalves in

Japan. *Phycological Research*, 43(3), pp.129–136. Available at: http://dx.doi.org/10.1111/j.1440-1835.1995.tb00016.x.

Hothorn, T., Bretz, F. & Westfall, P., 2016. *Simultaneous Inference in General Parametric Models, Package "multcomp,"* Available at: http://ftp5.gwdg.de/pub/misc/cran/web/packages/multcomp/multcomp.pdf.

Hunter, M.E. et al., 2015. Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons A. R. Mahon, ed. *PLoS ONE*, 10(4), p.e0121655.

Huver, J.R. et al., 2015. Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic ecosystems. *Ecological Applications*, 25(4), pp.991–1002.

J. Oksanen, F. Guillaume Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, H.W., 2016. Vegan: Community Ecology Package. R package. Available at: https://cran.r-project.org/package=vegan.

Jackson, J.B. et al., 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science (New York, N.Y.)*, 293(5530), pp.629–637.

Jane, S.F. et al., 2015. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology Resources*, 15(1), pp.216–227. Available at: http://doi.wiley.com/10.1111/1755-0998.12285.

Jennings, D.E. et al., 2012. Assessment of the aquatic biodiversity of a threatened coastal lagoon at Bimini, Bahamas. *Journal of Coastal Conservation*, 16(3), pp.405–428.

Jerde, C.L. et al., 2016. Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environmental Science and Technology*, 50(16), pp.8770–8779.

Jerde, C.L. et al., 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4(2), pp.150–157.

Ji, Y. et al., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16(10), pp.1245–1257.

Jo, T. et al., 2017. Rapid degradation of longer DNA fragments enables the improved

estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28449215 [Accessed April 28, 2017].

Juhel, J.-B. et al., 2017. Reef accessibility impairs the protection of sharks J. Blanchard, ed. *Journal of Applied Ecology*, p.n/a-n/a. Available at: http://dx.doi.org/10.1111/1365-2664.13007.

Kajiura, S.M. & Tellman, S.L., 2016. Quantification of massive seasonal aggregations of blacktip sharks (Carcharhinus limbatus) in southeast Florida. *PLoS ONE*, 11(3), pp.1–16.

Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772–780.

Kearse, M. et al., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Kelly, R., Gallego, R. & Jacobs-Palmer, E., 2017. The effect of tides on nearshore environmental DNA. Available at: https://peerj.com/preprints/3436/ [Accessed November 29, 2017].

Kelly, R.P. et al., 2017. Genetic and Manual Survey Methods Yield Different and Complementary Views of an Ecosystem. *Frontiers in Marine Science*, 3(January), pp.1–11. Available at: http://journal.frontiersin.org/article/10.3389/fmars.2016.00283/full.

Kelly, R.P. et al., 2016. Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, 4, p.e2444. Available at: https://doi.org/10.7717/peerj.2444.

Kelly, R.P., Port, J. a., et al., 2014. Harnessing DNA to improve environmental management. *Science*, 344(6191), pp.1455–1456.

Kelly, R.P., 2016. Making environmental DNA count. *Molecular Ecology Resources*, 16(1), pp.10–12. Available at: http://doi.wiley.com/10.1111/1755-0998.12455 [Accessed July 11, 2016].

Kelly, R.P., Port, J.A., et al., 2014. Using Environmental DNA to Census Marine

Fishes in a Large Mesocosm G. E. Hofmann, ed. *PLoS ONE*, 9(1), p.e86175. Available at: http://dx.plos.org/10.1371/journal.pone.0086175.

Kessel, S.T. et al., 2016. Three decades of longlining in Bimini, Bahamas, reveals long-term trends in lemon shark Negaprion brevirostris (Carcharhinidae) catch per unit effort. *Journal of Fish Biology*, 88(6), pp.2144–2156.

Kim, T.G., Jeong, S.-Y. & Cho, K.-S., 2014. Comparison of droplet digital PCR and quantitative real-time PCR for examining population dynamics of bacteria in soil. *Applied microbiology and biotechnology*, 98(13), pp.6105–6113.

Klobucar, S.L., Rodgers, T.W. & Budy, P., 2017. At the forefront: evidence of the applicability of using environmental DNA to quantify the abundance of fish populations in natural lentic waters with additional sampling considerations. *Canadian Journal of Fisheries and Aquatic Sciences*, (August), pp.1–5. Available at: http://www.nrcresearchpress.com/doi/10.1139/cjfas-2017-0114.

Klymus, K.E. et al., 2015. Quantification of eDNA shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp Hypophthalmichthys molitrix. *BIOLOGICAL CONSERVATION*, 183(SI), pp.77–84. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0006320714004455.

Klymus, K.E., Marshall, N.T. & Stepien, C.A., 2017. Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes H. Doi, ed. *PLoS ONE*, 12(5), p.e0177643. Available at: http://dx.plos.org/10.1371/journal.pone.0177643 [Accessed May 31, 2017].

Kreader, C.A., 1998. Persistence of PCR-Detectable Bacteroides distasonis from Human Feces in River Water Persistence of PCR-Detectable Bacteroides distasonis from Human Feces in River Water. , 64(10), pp.4103–4105.

Kulikova, T., 2004. The EMBL Nucleotide Sequence Database. *Nucleic Acids Research*, 32(90001), p.27D–30. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh120.

Lacoursière-Roussel, A. et al., 2016. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *Journal of Applied Ecology*, 53(4), pp.1148–1157. Available at: http://dx.doi.org/10.1111/1365-2664.12598.

Lacoursiere-Roussel, A., Rosabal, M. & Bernatchez, L., 2016. Estimating fish

abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources*, 16(6), pp.1401–1414.

Laramie, M.B., Pilliod, D.S. & Goldberg, C.S., 2015. Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biological Conservation*, 183, pp.29–37. Available at: http://www.sciencedirect.com/science/article/pii/S0006320714004509 [Accessed June 28, 2016].

Laszlo, A.H. et al., 2014. Decoding long nanopore sequencing reads of natural DNA. *Nature Biotechnology*, 32(8), pp.829–833.

Lawson Handley, L., 2015. How will the "molecular revolution" contribute to biological recording? *Biological Journal of the Linnean Society*, pp.750–766.

Lejzerowicz, F. et al., 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Scientific Reports*, 5(1), p.13932. Available at: http://www.nature.com/articles/srep13932.

Lennon, J.T., 2007. Diversity and metabolism of marine bacteria cultivated on dissolved DNA. *Applied and Environmental Microbiology*, 73(9), pp.2799–2805.

Leray, M. et al., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in zoology*, 10(1), p.34. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3686579&tool=pmcentrez&rendertype=abstract.

Leray, M. & Knowlton, N., 2016. Censusing marine eukaryotic diversity in the twenty-first century. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371(1702), p.20150331. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27481783%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4971183.

Leray, M. & Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, 2014(July), p.201424997. Available at:

http://www.pnas.org/lookup/doi/10.1073/pnas.1424997112.

Leray, M. & Knowlton, N., 2017. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ*, 5, p.e3006. Available at: https://peerj.com/articles/3006.

Lewis, R.J. et al., 2017. Applying the dark diversity concept to nature conservation. *Conservation Biology*, 31(1), pp.40–47.

Li, W. et al., 2010. Oomycetes and fungi: important parasites on marine algae. *Acta Oceanologica Sinica*, 29(5), pp.74–81. Available at: https://doi.org/10.1007/s13131-010-0065-4.

Liang, Z. & Keeley, A., 2013. Filtration recovery of extracellular DNA from environmental water samples. *Environmental Science and Technology*, 47(16), pp.9324–9331.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature*, 362, p.709. Available at: http://dx.doi.org/10.1038/362709a0.

Lindenmayer, D.B. & Likens, G.E., 2011. Direct Measurement Versus Surrogate Indicator Species for Evaluating Environmental Change and Biodiversity Loss. *Ecosystems*, 14(1), pp.47–59. Available at: https://doi.org/10.1007/s10021-010-9394-6.

Lindeque, P.K. et al., 2013. Next generation sequencing reveals the hidden diversity of zooplankton assemblages. *PLoS ONE*, 8(11), pp.1–14.

Lodge, D.M. et al., 2012. Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, 21(11), pp.2555–2558.

López-escardó, D. et al., 2018. Metabarcoding analysis reveals new metazoan diversity in European coastal samples. , 3.

Mace, G.M., Norris, K. & Fitter, A.H., 2012. Biodiversity and ecosystem services: a multilayered relationship. *Trends in Ecology & Evolution*, 27(1), pp.19–26. Available at: http://www.sciencedirect.com/science/article/pii/S0169534711002424.

Mackas, D.L. & Beaugrand, G., 2010. Comparisons of zooplankton time series. *Journal of Marine Systems*, 79(3), pp.286–304. Available at:

http://www.sciencedirect.com/science/article/pii/S0924796309000852.

Mahé, F. et al., 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3, p.e1420. Available at: https://peerj.com/articles/1420.

Mahon, A.R. et al., 2013. Validation of eDNA Surveillance Sensitivity for Detection of Asian Carps in Controlled and Field Experiments. *PLoS ONE*, 8(3), pp.1–6. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058316.

Maire, E. et al., 2016. How accessible are coral reefs to people? A global assessment based on travel time M. Holyoak, ed. *Ecology Letters*, 19(4), pp.351–360. Available at: http://doi.wiley.com/10.1111/ele.12577.

Markmann, M. & Tautz, D., 2005. Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp.1917–1924. Available at: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2005.1723.

Marschner, I., 2015. Fitting Generalized Linear Models, Package "glm2."

Martellini, A., Payment, P. & Villemur, R., 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. *Water Research*, 39(4), pp.541–548. Available at: http://www.sciencedirect.com/science/article/pii/S0043135404005378.

Martin, A.P., 1999. Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). *Molecular biology and evolution*, 16(7), pp.996–1002.

Martin, A.P., Naylor, G.J.P. & Palumbi, S.R., 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, 357, p.153. Available at: http://dx.doi.org/10.1038/357153a0.

Martin, a P., 1995. Mitochondrial DNA sequence evolution in sharks: rates, patterns, and phylogenetic inferences. *Molecular biology and evolution*, 12, pp.1114–1123.

Maruyama, A. et al., 2014. The Release Rate of Environmental DNA from Juvenile and Adult Fish M. Stöck, ed. *PLoS ONE*, 9(12), p.e114639. Available at:

http://dx.plos.org/10.1371/journal.pone.0114639 [Accessed July 11, 2016].

McClenachan, L., Cooper, A.B. & Dulvy, N.K., 2016. Rethinking Trade-Driven Extinction Risk in Marine and Terrestrial Megafauna. *Current Biology*, 26(12), pp.1640–1646. Available at: http://dx.doi.org/10.1016/j.cub.2016.05.026.

Mee, L.D., Espinosa, M. & Diaz, G., 1986. Paralytic shellfish poisoning with a Gymnodinium catenatum red tide on the Pacific Coast of Mexico. *Marine Environmental Research*, 19(1), pp.77–92. Available at: http://www.sciencedirect.com/science/article/pii/0141113686900401.

Meyer, W. & Seegers, U., 2012. Basics of skin structure and function in elasmobranchs: a review. *Journal of Fish Biology*, 80(5), pp.1940–1967. Available at: http://dx.doi.org/10.1111/j.1095-8649.2011.03207.x.

Minamoto, T. et al., 2017. Environmental DNA reflects spatial and temporal jellyfish distribution H. Doi, ed. *PLOS ONE*, 12(2), p.e0173073. Available at: http://dx.plos.org/10.1371/journal.pone.0173073 [Accessed March 17, 2017].

Minamoto, T. et al., 2012. Surveillance of fish species composition using environmental DNA. *Limnology*, 13(2), pp.193–197.

Di Minin, E. et al., 2013. Creating Larger and Better Connected Protected Areas Enhances the Persistence of Big Game Species in the Maputaland-Pondoland-Albany Biodiversity Hotspot. *PLoS ONE*, 8(8).

Miotke, L. et al., 2014. High sensitivity detection and quantitation of DNA copy number and single nucleotide variants with single color droplet digital PCR. *Analytical Chemistry*, 86(5), pp.2618–2624.

Miya, M. et al., 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, 2(7), p.150088. Available at: http://rsos.royalsocietypublishing.org/content/2/7/150088.

Moeslund, J.E. et al., 2017. Using dark diversity and plant characteristics to guide conservation and restoration. *Journal of Applied Ecology*, 54(6), pp.1730–1741. Available at: http://dx.doi.org/10.1111/1365-2664.12867.

Mora, C. et al., 2011. How many species are there on earth and in the ocean? *PLoS Biology*, 9(8), pp.1–8.

Morey-Gaines, G., 1982. Gymnodinium catenatum Graham (Dinophyceae): morphology and affinities with armoured forms. *Phycologia*, 21(2), pp.154–163. Available at: https://doi.org/10.2216/i0031-8884-21-2-154.1.

Mourier, J., Brown, C. & Planes, S., 2017. Learning and robustness to catch-and-release fishing in a shark social network. *Biology Letters*, 13(3), p.20160824. Available at: http://rsbl.royalsocietypublishing.org/lookup/doi/10.1098/rsbl.2016.0824.

Moyer, G.R. et al., 2014. Assessing environmental DNA detection in controlled lentic systems. *PLoS ONE*, 9(7).

Mullon, C., Freon, P. & Cury, P., 2005. The dynamcis of collapse in world fisheries. *Fish and Fisheries*, 6, pp.111–120.

Musick, J.A. et al., 2000. Management of Sharks and Their Relatives (Elasmobranchii). *Fisheries*, 25(3), pp.9–11.

Muter, B.A. et al., 2013. Australian and U.S. news media portrayal of sharks and their conservation. *Conservation Biology*, 27(1), pp.187–196.

Myers, R.A. & Worm, B., 2003. Rapid worldwide depletion of predatory fish communities. *Nature*, 423(6937), pp.280–283.

Nadon, M.O. et al., 2012. Re-Creating Missing Population Baselines for Pacific Reef Sharks. *Conservation Biology*, 26(3), pp.493–503.

Nathan, L.M. et al., 2014. Quantifying Environmental DNA Signals for Aquatic Invasive Species Across Multiple Detection Platforms. *Environmental Science & Technology*, 48(21), pp.12800–12806. Available at: http://pubs.acs.org/doi/abs/10.1021/es5034052.

Navia Andrés F., Mejía-Falla Paola A., López-García Juliana, Giraldo Alan, C.-E.V.H., 2017. How many trophic roles can elasmobranchs play in a marine tropical network? *Marine and Freshwater Research*, (January). Available at: http://www.publish.csiro.au/?paper=MF16161.

Newton, K. et al., 2007. Current and Future Sustainability of Island Coral Reef Fisheries. *Current Biology*, 17(7), pp.655–658.

Nielsen, K.M. et al., 2007. Release and persistence of extracellular DNA in the environment. *Environmental Biosafety Research*, 6(1–2), pp.37–53. Available at:

http://www.ebr-journal.org/10.1051/ebr:2007031.

O'Donnell, J.L. et al., 2017. Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ*, 5, p.e3044. Available at: https://peerj.com/articles/3044.

Ogram, A. et al., 1988. DNA adsorption to soils and sediments. *Environmental Science & Technology*, 22(8), pp.982–984. Available at: http://dx.doi.org/10.1021/es00173a020.

Ogram, A., Sayler, G.S. & Barkay, T., 1987. The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods*, 7(2–3), pp.57–66.

Olds, B.P. et al., 2016. Estimating species richness using environmental DNA. *Ecology and Evolution*, 6(12), pp.4214–4226.

Olson, Z.H., Briggler, J.T. & Williams, R.N., 2012. An eDNA approach to detect eastern hellbenders (Cryptobranchus a. alleganiensis) using samples of water. *Wildlife Research*, 39(7), pp.629–636.

Ottesen, E.A., 2016. Probing the living ocean with ecogenomic sensors. *Current Opinion in Microbiology*, 31, pp.132–139. Available at: http://dx.doi.org/10.1016/j.mib.2016.03.012.

Palumbo, J.D., O'Keeffe, T.L. & Fidelibus, M.W., 2016. Characterization of Aspergillus section Nigri species populations in vineyard soil using droplet digital PCR. *Letters in applied microbiology*, 63(6), pp.458–465.

Pärtel, M., Szava-Kovats, R. & Zobel, M., 2013. Community Completeness: Linking Local and Dark Diversity within the Species Pool Concept. *Folia Geobotanica*, 48(3), pp.307–317. Available at: https://doi.org/10.1007/s12224-013-9169-x.

Pärtel, M., Szava-Kovats, R. & Zobel, M., 2018. Dark diversity: shedding light on absent species. *Trends in Ecology & Evolution*, 26(3), pp.124–128. Available at: http://dx.doi.org/10.1016/j.tree.2010.12.004.

Paul, J.H., Kellogg, C.A. & Jiang, S.C., 1996. Viruses and DNA in Marine Environments. In R. R. Colwell, U. Simidu, & K. Ohwada, eds. *Microbial Diversity in Time and Space*. Boston, MA: Springer US, pp. 115–124. Available at: https://doi.org/10.1007/978-0-585-34046-3_14.

Pawlowski, J. et al., 2014. Environmental monitoring through protist next-generation

sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources*, 14(6), pp.1129–1140. Available at: http://dx.doi.org/10.1111/1755-0998.12261.

Piaggio, A.J. et al., 2014. Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Molecular Ecology Resources*, 14(2), pp.374–380. Available at: http://doi.wiley.com/10.1111/1755-0998.12180.

Pilliod, D.S. et al., 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(8), pp.1123–1130. Available at: http://www.nrcresearchpress.com/doi/abs/10.1139/cjfas-2013-0047#.U5-Jn9Eg_IU.

Pilliod, D.S. et al., 2014. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, 14(1), pp.109–116.

Pilskaln, C.H. et al., 2005. High concentrations of marine snow and diatom algal mats in the North Pacific Subtropical Gyre: Implications for carbon and nitrogen cycles in the oligotrophic ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 52(12), pp.2315–2332. Available at: http://www.sciencedirect.com/science/article/pii/S0967063705002050.

Pindera, A.C., Gozlan, R.E. & Britton, J.R., 2005. Dispersal of the invasive topmouth gudgeon, Pseudorasbora parva in the UK: A vector for an emergent infectious disease. *Fisheries Management and Ecology*, 12(6), pp.411–414.

Pinheiro, L.B. et al., 2012. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Analytical Chemistry*, 84(2), pp.1003–1011.

Pinol, J. et al., 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular ecology resources*, 15(4), pp.819–830.

Pochon, X. et al., 2017. Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications H. Doi, ed. *PLOS ONE*, 12(11), p.e0187636. Available at:

http://dx.plos.org/10.1371/journal.pone.0187636 [Accessed November 3, 2017].

Pompanon, F. et al., 2012. Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology*, 21(8), pp.1931–1950.

Port, J.A. et al., 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, 25(2), pp.527–541.

Preston, C.M. et al., 2011. Underwater Application of Quantitative PCR on an Ocean Mooring. *PLoS ONE*, 6(8), p.e22522. Available at: http://dx.plos.org/10.1371/journal.pone.0022522.

Ratnasingham, S. & Hebert, P.D.N., 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7(3), pp.355–364. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1890991/.

Rees, H.C. et al., 2014. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51(5), pp.1450–1459.

Renshaw, M.A. et al., 2015. The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, 15(1), pp.168–176.

Riaz, T. et al., 2011. EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, 39(21), pp.1–11.

Robbins, W.D. et al., 2006. Ongoing Collapse of Coral-Reef Shark Populations. *Current Biology*, 16(23), pp.2314–2319.

Robson, H.L.A. et al., 2016. Fine-tuning for the tropics: application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Molecular Ecology Resources*, 16(4), pp.922–932. Available at: http://doi.wiley.com/10.1111/1755-0998.12505 [Accessed July 11, 2016].

Rognes, T. et al., 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, 4, p.e2584. Available at: https://peerj.com/articles/2584.

Romanowski, G. et al., 1992. Persistence of free plasmid DNA in soil monitored by various methods, including a transformation assay. *Applied and Environmental Microbiology*, 58(9), pp.3012–3019.

Sandin, S.A. et al., 2008. Baselines and Degradation of Coral Reefs in the Northern

Line Islands N. Ahmed, ed. *PLoS ONE*, 3(2), p.e1548. Available at: http://dx.plos.org/10.1371/journal.pone.0001548.

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), pp.5463–5467. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.74.12.5463.

Sassoubre, L.M. et al., 2016. Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environmental Science and Technology*, 50(19), pp.10456–10464.

Schmidt, B.R. et al., 2013. Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, 4(7), pp.646–653.

Scholin, C.A., 2010. What are " ecogenomic sensors? " A review and thoughts for the future. *Ocean Sci*, 6, pp.51–60. Available at: www.ocean-sci.net/6/51/2010/.

Secondi, J. et al., 2016. Detection of a global aquatic invasive amphibian, Xenopus laevis, using environmental DNA. *Amphibia Reptilia*, 37(1), pp.131–136.

Seddon, P.J. et al., 2014. Reversing defaunation: Restoring species in a changing world. *Science*, 345(6195), pp.406–412. Available at: http://science.sciencemag.org/content/345/6195/406.

Sharma, P. & Kobayashi, T., 2014. Are "universal" dna primers really universal? *Journal of Applied Genetics*, 55(4), pp.485–496.

Shaw, J.L.A. et al., 2016. Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, 197, pp.131–138. Available at: http://linkinghub.elsevier.com/retrieve/pii/S000632071630088X [Accessed July 11, 2016].

Shelton, A.O. et al., 2016. A framework for inferring biological communities from environmental DNA. *Ecological Applications*. Available at: http://doi.wiley.com/10.1890/15-1733.1 [Accessed July 11, 2016].

Shogren, A.J. et al., 2016. Modelling the transport of environmental DNA through a porous substrate using continuous flow-through column experiments. *Journal of*

The Royal Society Interface, 13(119), p.20160290. Available at: http://rsif.royalsocietypublishing.org/lookup/doi/10.1098/rsif.2016.0290.

Sigsgaard, E.E. et al., 2015. Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *BIOLOGICAL CONSERVATION*, 183(SI), pp.46–52. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0006320714004480.

Sigsgaard, E.E. et al., 2016. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Publishing Group*, 1(November), pp.1–4. Available at: http://dx.doi.org/10.1038/s41559-016-0004.

Sigsgaard, E.E. et al., 2017. Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, 164(6), p.128. Available at: http://link.springer.com/10.1007/s00227-017-3147-4 [Accessed May 16, 2017].

Simpfendorfer, C.A. et al., 2016. Environmental DNA detects Critically Endangered largetooth sawfish in the wild. *Endangered Species Research*, 30, pp.109–116. Available at: http://www.int-res.com/abstracts/esr/v30/p109-116/.

Simpfendorfer, C.A. et al., 2002. Results of a fishery-independent survey for pelagic sharks in the western North Atlantic, 1977-1994. *Fisheries Research*, 55(1–3), pp.175–192.

Simpfendorfer, C.A. et al., 2011. The importance of research and public opinion to conservation management of sharks and rays: A synthesis. *Marine and Freshwater Research*, 62(6), pp.518–527.

Sims, D.W. et al., 2006. Hunt warm, rest cool: Bioenergetic strategy underlying diel vertical migration of a benthic shark. *Journal of Animal Ecology*, 75(1), pp.176–190.

Sims, D.W. et al., 2004. Low-temperature-driven early spawning migration of a temperate marine fish. *Journal of Animal Ecology*, 73(2), pp.333–341.

Singler, H.R. & Villareal, T.A., 2005. Nitrogen inputs into the euphotic zone by vertically migrating Rhizosolenia mats. *Journal of Plankton Research*, 27(6), pp.545–556.

Snelgrove, P.V.R., 1999. Getting to the Bottom of Marine Biodiversity: Sedimentary

Habitats Ocean bottoms are the most widespread habitat on Earth and support high biodiversity and key ecosystem services. *BioScience*, 49(2), pp.129–138.

Soliveres, S. et al., 2016. Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature*, 536(7617), pp.456–459. Available at: http://dx.doi.org/10.1038/nature19092.

Sorenson, L., Santini, F. & Alfaro, M.E., 2014. The effect of habitat on modern shark diversification. *Journal of Evolutionary Biology*, 27(8), pp.1536–1548.

de Souza, L.S. et al., 2016. Environmental DNA (eDNA) Detection Probability Is Influenced by Seasonal Activity of Organisms H. Doi, ed. *PLOS ONE*, 11(10), p.e0165273. Available at: http://dx.plos.org/10.1371/journal.pone.0165273 [Accessed November 4, 2016].

Spaet, J.L.Y. & Berumen, M.L., 2015. Fish market surveys indicate unsustainable elasmobranch fisheries in the Saudi Arabian Red Sea. *Fisheries Research*, 161, pp.356–364. Available at: http://dx.doi.org/10.1016/j.fishres.2014.08.022.

Spear, S.F. et al., 2015. Using environmental DNA methods to improve detectability in a hellbender (Cryptobranchus alleganiensis) monitoring program. *BIOLOGICAL CONSERVATION*, 183(SI), pp.38–45. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0006320714004418.

Spens, J. et al., 2017. Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*, 8(5), pp.635–645.

Stat, M. et al., 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, 7(1), p.12240. Available at: http://www.nature.com/articles/s41598-017-12501-5.

Stevens, J., 2000. The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems. *ICES Journal of Marine Science*, 57(3), pp.476–494. Available at: http://icesjms.oxfordjournals.org/cgi/doi/10.1006/jmsc.2000.0724.

Stoeckle, M.Y., Soboleva, L. & Charlop-Powers, Z., 2017. Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary H. Doi, ed. *PLoS ONE*, 12(4), p.e0175186. Available at:

http://dx.plos.org/10.1371/journal.pone.0175186 [Accessed April 14, 2017].

Strickler, K.M., Fremier, A.K. & Goldberg, C.S., 2015. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, 183, pp.85–92. Available at: http://dx.doi.org/10.1016/j.biocon.2014.11.038.

T. Maniatis, E.F. Fritsch, J.S., 1982. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

Taberlet, P. et al., 2012. Environmental DNA. *Molecular Ecology*, 21, pp.1789–1793. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05542.x/full%5Cnpapers2://publication/uuid/3E2B6116-3B64-435F-A8EF-EFD5323DFD46.

Takahara, T. et al., 2012. Estimation of fish biomass using environmental DNA. *PLoS ONE*, 7(4), pp.3–10.

Takahara, T., Minamoto, T. & Doi, H., 2013. Using Environmental DNA to Estimate the Distribution of an Invasive Fish Species in Ponds S. Consuegra, ed. *PLoS ONE*, 8(2), p.e56584. Available at: http://dx.plos.org/10.1371/journal.pone.0056584.

Tang, C.Q. et al., 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences*, 109(40), pp.16208–16212. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1209160109.

Thomsen, P.F., Kielgast, J., Iversen, L.L., et al., 2012. Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples S. Lin, ed. *PLoS ONE*, 7(8), p.e41732. Available at: http://dx.plos.org/10.1371/journal.pone.0041732.

Thomsen, P.F. et al., 2016. Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes A. R. Mahon, ed. *PLOS ONE*, 11(11), p.e0165252. Available at: https://doi.org/10.1371/journal.pone.0165252.

Thomsen, P.F., Kielgast, J., Iversen, L.L., et al., 2012. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21(11),

pp.2565–2573. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22151771.

Thomsen, P.F. & Willerslev, E., 2015. Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, pp.4–18. Available at: http://dx.doi.org/10.1016/j.biocon.2014.11.019.

Tirard, P., 2011. *Requins du caillou*,

Tomlinson, J.A. et al., 2005. On-Site DNA Extraction and Real-Time PCR for Detection of Phytophthora ramorum in the Field On-Site DNA Extraction and Real-Time PCR for Detection of Phytophthora ramorum in the Field. *Applied and Environmental Microbiology*, 71(11), pp.6702–6710.

Torti, A., Lever, M.A. & Jørgensen, B.B., 2015. Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, 24, pp.185–196. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1874778715300234.

Tsutsui, S. et al., 2009. Common Skate (Raja kenojei) Secretes Pentraxin into the Cutaneous Secretion: The First Skin Mucus Lectin in Cartilaginous Fish*. *The Journal of Biochemistry*, 146(2), pp.295–306. Available at: http://dx.doi.org/10.1093/jb/mvp069.

Turner, C.R., Miller, D.J., et al., 2014. Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (hypophthalmichthys spp.). *PLoS ONE*, 9(12), p.e114329. Available at: http://dx.plos.org/10.1371/journal.pone.0114329.

Turner, C.R., Barnes, M.A., et al., 2014. Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, 5(7), pp.676–684.

Turner, C.R., Uy, K.L. & Everhart, R.C., 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, 183, pp.93–102. Available at: http://dx.doi.org/10.1016/j.biocon.2014.11.017.

Uchii, K., Doi, H. & Minamoto, T., 2016. A novel environmental DNA approach to quantify the cryptic invasion of non-native genotypes. *Molecular Ecology Resources*, 16(2), pp.415–422.

Untergasser, A. et al., 2012. Primer3-new capabilities and interfaces. *Nucleic Acids*

Research, 40(15), pp.1–12.

Ussler, W. et al., 2013. Autonomous application of quantitative PCR in the deep sea: In Situ surveys of aerobic methanotrophs using the deep-sea environmental sample processor. *Environmental Science and Technology*, 47(16), pp.9339–9346.

Valentini, A. et al., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25(4), pp.929–942.

Vamos, E., Elbrecht, V. & Leese, F., 2017. Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics*, 1, p.e14625. Available at: https://mbmg.pensoft.net/articles.php?id=14625.

de Vargas, C. et al., 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348(6237), pp.1261605–1261605. Available at: http://www.sciencemag.org.proxy.libraries.rutgers.edu/content/348/6237/1261605.full%5Cnhttp://www.sciencemag.org/cgi/doi/10.1126/science.1261605.

Vaudo, J.J. & Heithaus, M.R., 2012. Diel and seasonal variation in the use of a nearshore sandflat by a ray community in a near pristine system. *Marine and Freshwater Research*, 63(11), pp.1077–1084. Available at: https://doi.org/10.1071/MF11226.

Vaulot, D., Romari, K. & Not, F., 2002. Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends in microbiology*, 10(6), pp.266–267.

Venables, W.N. & Ripley, B.D., 2010. *Modern Applied Statistics with S*, Springer Publishing Company, Incorporated.

Vestheim, H. & Jarman, S.N., 2008. Blocking primers to enhance PCR amplification of rare sequences in mixed samples - a case study on prey DNA in Antarctic krill stomachs. *Frontiers in zoology*, 5, p.12. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2517594&tool=pmcentrez&rendertype=abstract.

Villarino, E. et al., 2018. Large-scale ocean connectivity and planktonic body size. *Nature Communications*, 9(1), p.142. Available at: http://www.nature.com/articles/s41467-017-02535-8.

Vogelstein, B. & Kinzler, K.W., 1999. Digital PCR. *Genetics*, 96(August), pp.9236–

9241.

Wangensteen, Palacín, Guardiola, T., 2017. Metabarcoding littoral hard-bottom communities: unexpected diversity and database gaps revealed by two molecular markers. *PeerJ Preprints*.

Wangensteen, O.S.. et al., 2017. Metabarcoding shallow marine hard-bottom communities: unexpected diversity and database gaps revealed by two molecular markers. *Molecular Ecology*.

Wangensteen, O.S., Cebrian, E. & Palacín, C., 2018. Under the canopy: Community-wide effects of invasive algae in Marine Protected Areas revealed by metabarcoding. *Marine Pollution Bulletin*, 127(November 2017), pp.54–66. Available at: https://doi.org/10.1016/j.marpolbul.2017.11.033.

Wangensteen, O.S. & Turon, X., 2017. Metabarcoding Techniques for Assessing Biodiversity of Marine Animal Forests. In C. Editors: Rossi, S., Bramanti, L., Gori, A., Orejas, ed. *Marine Animal Forests*. Cham: Springer International Publishing, pp. 1–29. Available at: http://link.springer.com/10.1007/978-3-319-17001-5_53-1.

Ward-Paige, C.A. et al., 2010. Large-scale absence of sharks on reefs in the greater-caribbean: A footprint of human pressures. *PLoS ONE*, 5(8).

Ward, R.D. et al., 2005. DNA barcoding Australia's fish species. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(September), pp.1847–1857.

Webb, T.J. & Mindel, B.L., 2015. Global patterns of extinction risk in marine and non-marine systems. *Current Biology*, 25(4), pp.506–511. Available at: http://dx.doi.org/10.1016/j.cub.2014.12.023.

Weinbauer, M.G., Fuks, D. & Peduzzi, P., 1993. Distribution of viruses and dissolved DNA along a coastal trophic gradient in the northern Adriatic Sea. *Applied and Environmental Microbiology*, 59(12), pp.4074–4082.

Weise, F.J., Stratford, K.J. & Van Vuuren, R.J., 2014. Financial costs of large carnivore translocations - Accounting for conservation. *PLoS ONE*, 9(8).

Weltz, K. et al., 2017. Application of environmental DNA to detect an endangered marine skate species in the wild. *PLoS ONE*, 12(6), pp.1–16.

Werry, J.M. et al., 2014. Reef-Fidelity and Migration of Tiger Sharks, Galeocerdo cuvier, across the Coral Sea A. P. Klimley, ed. *PLoS ONE*, 9(1), p.e83249. Available at: http://dx.plos.org/10.1371/journal.pone.0083249.

Wheeler, Q.D., 2004. Taxonomy: Impediment or Expedient? *Science (New York, N.Y.)*, 303(5656), pp.285–285.

Wilcox, T.M. et al., 2013. Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity R. C. Willson, ed. *PLoS ONE*, 8(3), p.e59520. Available at: http://dx.plos.org/10.1371/journal.pone.0059520.

Wilcox, T.M. et al., 2016. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. *Biological Conservation*, 194, pp.209–216. Available at: http://www.sciencedirect.com/science/article/pii/S0006320715302020 [Accessed December 31, 2015].

Willerslev, E. et al., 2009. Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. , 317(5834), pp.111–114.

Willerslev, E. et al., 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, 300. Available at: http://dx.doi.org/10.1126/science.1084114.

Williams, K.E., Huyvaert, K.P. & Piaggio, A.J., 2016. No filters, no fridges: a method for preservation of water samples for eDNA analysis. *BMC research notes*, 9(1), p.298. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27278936 [Accessed June 29, 2016].

Wilson, C.C., Wozney, K.M. & Smith, C.M., 2015. Recognizing false positives: Synthetic oligonucleotide controls for environmental DNA surveillance. *Methods in Ecology and Evolution*.

Wilson, I.G., 1997. Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology*, 63(10), pp.3741–3751.

Worm, B. et al., 2013. Global catches, exploitation rates, and rebuilding options for sharks. *Marine Policy*, 40(1), pp.194–204. Available at: http://dx.doi.org/10.1016/j.marpol.2012.12.034.

Yamamoto, S. et al., 2016. Environmental DNA as a “snapshot” of fish distribution: A

case study of Japanese jack mackerel in Maizuru Bay, Sea of Japan. *PLoS ONE*, 11(3), pp.1–18.

Yamamoto, S. et al., 2017. Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports*, 7(January), p.40368. Available at: http://www.nature.com/articles/srep40368.

Yamanaka, H. et al., 2016. On-site filtration of water samples for environmental DNA analysis to avoid DNA degradation during transportation. *Ecological Research*, 31(6), pp.963–967. Available at: https://doi.org/10.1007/s11284-016-1400-9.

Yang, L. et al., 2014. Species identification through mitochondrial rRNA genetic analysis. *Scientific Reports*, 4, pp.1–11.

Young, H.S. et al., 2016. Patterns, Causes, and Consequences of Anthropocene Defaunation. *Annual Review of Ecology, Evolution, and Systematics*, 47(1), pp.333–358. Available at: http://www.annualreviews.org/doi/10.1146/annurev-ecolsys-112414-054142.

Zaiko, A. et al., 2015. Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Marine Pollution Bulletin*, 100(1), pp.53–59. Available at: http://dx.doi.org/10.1016/j.marpolbul.2015.09.030.

Zeale, M.R.K. et al., 2011. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, 11(2), pp.236–244. Available at: http://dx.doi.org/10.1111/j.1755-0998.2010.02920.x.

Zhang, Y. et al., 2013. Intracellular and Extracellular Antimicrobial Resistance Genes in the Sludge of Livestock Waste Management Structures. *Environmental Science & Technology*, 47(18), pp.10206–10213. Available at: http://dx.doi.org/10.1021/es401964s.

Zhou, X. et al., 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2(1), pp.1–12.